



**PCT**

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|  |                  |  |
|--|------------------|--|
| <p>(51) International Patent Classification <sup>6</sup> :<br/><b>G06F 17/00, 159/00</b></p>   | <p><b>A1</b></p> | <p>(11) International Publication Number: <b>WO 96/34347</b><br/>(43) International Publication Date: 31 October 1996 (31.10.96)</p>   |
| <p>(21) International Application Number: PCT/US96/06110<br/>(22) International Filing Date: 26 April 1996 (26.04.96)<br/>(30) Priority Data:<br/>08/431,155 27 April 1995 (27.04.95) US<br/>(60) Parent Application or Grant<br/>(63) Related by Continuation<br/>US Not furnished (CIP)<br/>Filed on Not furnished<br/>(71) Applicant (for all designated States except US): MOUNT SINAI SCHOOL OF MEDICINE OF THE CITY UNIVERSITY OF NEW YORK [US/US]; One Gustave L. Levy Place, New York, NY 10029-6574 (US).<br/>(72) Inventor; and<br/>(75) Inventor/Applicant (for US only): GUARNIERI, Frank [US/US]; 62 Lake Street, Brooklyn, NY 11223 (US).<br/>(74) Agents: CLARK, Richard, S. et al.; Brumbaugh, Graves, Donohue &amp; Raymond, 30 Rockefeller Plaza, New York, NY 10112 (US).</p> |                  | <p>(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).<br/><br/><b>Published</b><br/><i>With international search report.<br/>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p> |
| <p>(54) Title: METHOD FOR IDENTIFYING STRUCTURALLY ACTIVE COMPOUNDS USING CONFORMATIONAL MEMORIES<br/>(57) Abstract<br/><br/>The present invention relates to a method for predicting the conformation and functionality of a molecule, comprising the steps of, first, performing multiple simulated annealing runs in order to reveal populated and unpopulated regions of multidimensional conformation space, and, second, performing a simulation at a fixed temperature, with sampling only from populated regions found in the first step.</p>  |                  |  |

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |                                       |    |                          |
|----|--------------------------|----|---------------------------------------|----|--------------------------|
| AM | Armenia                  | GB | United Kingdom                        | MW | Malawi                   |
| AT | Austria                  | GE | Georgia                               | MX | Mexico                   |
| AU | Australia                | GN | Guinea                                | NE | Niger                    |
| BB | Barbados                 | GR | Greece                                | NL | Netherlands              |
| BE | Belgium                  | HU | Hungary                               | NO | Norway                   |
| BF | Burkina Faso             | IE | Ireland                               | NZ | New Zealand              |
| BG | Bulgaria                 | IT | Italy                                 | PL | Poland                   |
| BJ | Benin                    | JP | Japan                                 | PT | Portugal                 |
| BR | Brazil                   | KE | Kenya                                 | RO | Romania                  |
| BY | Belarus                  | KG | Kyrgyzstan                            | RU | Russian Federation       |
| CA | Canada                   | KP | Democratic People's Republic of Korea | SD | Sudan                    |
| CF | Central African Republic | KR | Republic of Korea                     | SE | Sweden                   |
| CG | Congo                    | KZ | Kazakhstan                            | SG | Singapore                |
| CH | Switzerland              | LI | Liechtenstein                         | SI | Slovenia                 |
| CI | Côte d'Ivoire            | LK | Sri Lanka                             | SK | Slovakia                 |
| CM | Cameroon                 | LR | Liberia                               | SN | Senegal                  |
| CN | China                    | LT | Lithuania                             | SZ | Swaziland                |
| CS | Czechoslovakia           | LU | Luxembourg                            | TD | Chad                     |
| CZ | Czech Republic           | LV | Latvia                                | TG | Togo                     |
| DE | Germany                  | MC | Monaco                                | TJ | Tajikistan               |
| DK | Denmark                  | MD | Republic of Moldova                   | TT | Trinidad and Tobago      |
| EE | Estonia                  | MG | Madagascar                            | UA | Ukraine                  |
| ES | Spain                    | ML | Mali                                  | UG | Uganda                   |
| FI | Finland                  | MN | Mongolia                              | US | United States of America |
| FR | France                   | MR | Mauritania                            | UZ | Uzbekistan               |
| GA | Gabon                    |    |                                       | VN | Viet Nam                 |

### Description

#### METHOD FOR IDENTIFYING STRUCTURALLY ACTIVE COMPOUNDS USING CONFORMATIONAL MEMORIES

##### Introduction

The present invention relates to a method for predicting the conformation and functionality of a molecule, comprising the steps of, first, performing multiple simulated annealing runs in order to reveal populated and unpopulated regions of multidimensional conformation space, and, second, performing a simulation at a fixed temperature, with sampling only from populated regions found in the first step.

##### Background of the Invention

The insights gained from simulations, and the growing prevalence of relatively inexpensive computer power, has led to the widespread use of many computational techniques. The successes of these methods has prompted the continuing development of new methods to study more and more complex problems. Early chemical simulations, for example, were used to estimate equilibrium statistical mechanical quantities and transport properties on collections of point particles whose interactions were governed by simple potentials (Alder and Wainwright, Phys. Rev. Lett., 18:988 (1967); Hoover and Ree, J. Chem Phys. 49:3609 (1968); Rahman, Phys. Rev. 136:A405 (1964); Verlet, Phys. Rev. 159:98 (1967)). Today, simulations are used to compute free energies (Jorgensen, Acc. Chem. Res. 22:184 (1989); Kollman, Chem. Rev. 93:2395 (1993)) and to study complex systems like protein structure (McCammon and Harvey, in Dynamics of Proteins and Nucleic Acids, Cambridge University Press, New York, N.Y. (1987)). Increasing demands naturally lead to increasing difficulties. One of the great difficulties in compu-

tational chemistry is the simulation of flexible organic and biological molecules. These systems are problematic because they belong to the general class of problems known as multiple time scale problems

5 (Brackbill and Cohen, in Multiple Time Scales, Academic Press, Orlando Florida, (1985)). Flexible organic molecules belong to this class because bond length and bond angle motion occurs on a femtosecond to picosecond time scale, while torsional motion may occur on a nano-  
10 second time scale or longer. Since true convergence of statistical mechanical properties requires multiple interconversions between all torsional states, obtaining stable averages may require simulations on the order of tens or even hundreds of nanoseconds.

15 Typically, chemical systems are simulated using Monte Carlo ("MC" Metropolis et al., J. Chem Phys. 21:187 (1953)) or molecular dynamics ("MD"; Allen and Tildesley, in Computer Simulations of Liquids, (Clarendon, Oxford, (1987)). The recognition that the  
20 study of complex systems may overtax these standard methods has led to much work on the development of more powerful techniques. Several different variations and extensions of these two basic procedures have been tried in an attempt to find more efficient methods.

25 These new algorithms generally fall into two broad classes: MD methods with the addition of some random character, (Anderson, J. Chem Phys. 72:2384 (1980); van Gunsteren and Berendsen, Mol. Sim., 1:173 (1988)) and MC methods which utilize some partial deterministic  
30 character to generate better trial moves (Brass et al., Biopol., 33:1307 (1993); Heerman et al., Comp. Phys Comm., 60:311 (1990); Cao and Berne, J. Chem Phys., 92:1980 (1990); Rao and Berne, J. Chem Phys., 71:129 (1979); Rossky et al., J. Chem Phys., 69:4628 (1978)).

35 The most recent algorithm, the mixed MC-SD algorithm,

is a pure hybrid that uses MD and MC methods equally (Burger et al., J. Amer. Chem. Soc., 116:8, 3593).

In the MC method, increases in efficiency may be obtained by generating nonuniform random trial moves. If this is done over the same surface being studied, it is known as biased sampling. If the searching is carried out over a potential surface that is different from the surface being studied, it is characterized as importance sampling. Since importance sampling is done over a different surface than the one actually being studied, it is necessary to appropriately weight the statistical mechanical averages (Kalos and Whitlock, Monte Carlo Methods vol 1, (Wiley, New York, N.Y. 1986)). In biased sampling, since the searching is being done over the same potential surface as the actual surface under study, no weighting of the averages is necessary. In biased sampling, MC trial moves are based on some a priori knowledge of the space to be sampled. Regions that are more "important" are sampled with greater frequency in the full expectation that the speed of convergence will be enhanced. In practice, a simulation employing biased sampling is done in two steps: some initial procedure is employed to reveal (or guess) the important gross features, then an extensive search utilizing this information is performed. Obviously, this procedure will be valid and useful only if it is faster than a standard sampling procedure, and if it introduces no spurious artifacts.

### 30 SUMMARY OF THE INVENTION

The present invention relates to a method for identifying structurally active molecules comprising the steps of, first, performing multiple simulated annealing runs in order to reveal populated and unpopulated regions of multidimensional conformation space, and, second, performing a simulation at a fixed tem-

perature, with sampling only from populated regions found in the first step. It is based, at least in part, on the discovery that the method of the invention could be used to sort a large family of analogs of gonadotrophin releasing hormone ("GnRH analogs") into groups having low or high affinity for the GnRH receptor. The method of the present invention offers the advantage that, since the simulated annealing runs quickly reveal unpopulated regions of the conformation space, the volume of conformation space that needs to be sampled in the second phase of the algorithm is reduced by many orders of magnitude. Additionally, since no energy minimization is used, these populations represent a canonical ensemble which may be used to estimate conformational free energies.

#### Description of the Figures

Figure 1. Structure of  $\text{LTB}_4$ .

Figure 2. A Flex-Map or "Conformational Memory" of dihedral 1 from  $\text{LTB}_4$ .

Figure 3. Graphical Representation of the Mapping of Rand Numbers onto the dihedral distribution. Random numbers between 0-1 determine dihedral values (using the ---- line). For example, 0.6 maps to  $+65^\circ$ .

Figure 4. Plot of the average  $\text{LTB}_4$  conformer energy vs temperature for ten normal SA runs and ten Smart-SA runs. Very rapid energy lowering is possible using Smart-SA, although the ultimate energy is similar. Each 10 run set required ~16 hrs of CPU time on a Vax 8600 and represents 100,000 conformers each.

Figure 5. Molecular structure of the gonadotropin-releasing hormone (GnRH). The 35 rotatable torsional angles are indicated by arrows.

Figure 6. Conformational Memories of selected dihedral angles in the gonadotropin-releasing hormone (see Fig. 5 for identity of the angles. a. Dihedral

angle 4; b. Dihedral angle 6; c. Dihedral angle 7; d. Dihedral angle 35.

Figure 7. Conformational Memory difference maps of dihedral angle 20 in GnRH (see Fig. 5). The difference maps were created by subtracting the Conformational Memories from A. 25 and 10 runs, B. 50 and 25 runs, C. 75 and 50 runs, D. 100 and 75 runs. Note that in almost all regions differences are <1%. Conformational Memory Difference Maps of the other dihedrals are very similar.

Figure 8. A sequences of 310K temperature slices from the Conformational Memory of bond 17, calculated with a. 25 runs; b. 50 runs; c. 75 runs; d. 100 runs. Note the symmetrically equivalent population distributions centered about -90 and 90.

Figure 9. The choice of dihedral angle values in the biased sampling of the populated region of the Conformational Memories. The illustration is for dihedral angle 19. Panel (a) shows a histogram representation of the probability distribution for the dihedral angle; panel (b) shows the cumulative probability distribution for dihedral angle 19. Since the random number generator is a cumulative probability distribution, biased sampling is done from the histogram in part (b). If the random number 0.2 is generated, which corresponds to the second block of the histogram in part (b), the new trial dihedral will be chosen from the interval -170 to -160 degrees with the actual value obtained from a linear interpolation within this interval. If the random number 0.4 is generated, which corresponds to the 28th block of the histogram in part (b), the new trial dihedral will be chosen from the interval 90 to 100 degrees. Note that the region -60 and 60, which has no population in part (a), is automatically skipped when sampling from part (b).

Figure 10. Backbone trace of a representative of the five conformational families of GnRH obtained from Conformational Memories. Structures with a beta-type turn have a 70% population. Structures with a straight backbone have approximately 5% population.

Figure 11. Superimposition of 70 structures that make up the major conformational family of GnRH obtained from Conformational memories. While there is a large amount of fluctuation in the backbone, and an even greater amount of fluctuation in the side chains (especially Arg8), there is a clear beta-type turn from residues 5-8 in this family.

Figure 12. Backbone trace of a representative of the two conformational families of Lys8-GnRH obtained from Conformational Memories. The structure with the beta-type turn comes from a family with an approximately 3% population. The structure with the straight backbone comes from a family with approximately that of 70% population.

Figure 13. Superimposition of 70 structures that make up the major conformational family of Lys8-GnRH obtained from Conformational Memories. While there is a large amount of fluctuation in the backbone, and an even greater amount of fluctuation in the side chains (especially Lys8), the backbone is clearly extended.

Figure 14. Superimposition of a high affinity GnRH cyclic analog (Structure I) and a representative of the major GnRH conformational family (structure II). Eleven backbone atoms from residues 5-8 were used for the superimposition.

#### Detailed Description of the Invention

For purposes of clarity of description, and not by way of limitation, the present invention is described by way of two examples. First, the method of the invention is applied to determining the structure of leukotrienes. Second, the method of the invention is



used to identify GnRH analogs which have a high affinity of binding to the GnRH receptor.

Example: Determination of Leukotriene Structure

5       Leukotrienes, for example, are an important class of natural antiinflammatory agents (Sammuelsson et al., Prostaglandins, 17:785 (1979); "The Leukotrienes": Their Biological Significance," P.J. Piper, Ed., Raven Press, N.Y., (1986)). Understanding the bioactive con-  
10       formations of a key member of this class such as LTB<sub>4</sub> (Figure 1), involves conformational analysis of 14 flexible dihedrals.

      This is, however, an extremely difficult problem. For a description of "Impossible" computational pro-  
15       blems see: W. Garey, Computers and Intractability, (H. Freeman and Co., New York, N.Y. 1979). Even con- sidering only a three state model around each bond (anti and +/- gauche) there are 3<sup>14</sup> possible con- formations (Kirkpatrick, et al., Science, 220:671  
20       (1983); Simulated Annealing and Optimization, M.W. Johnson, Ed., American Sciences Press, Syracuse, N.Y. (1988)) (1,594,323). A recent case study on the conformational analysis of cycloheptadecane (Saunders et al., J. Amer. Chem. Soc. 112:1419 (1990)) which  
25       state is equivalent to a 12-dimensional nonsymmetric problem, nicely illustrates the difficulties of searching a multidimensional conformational space.

Smart-Simulated Annealing: The Learning Phase

30       To address this problem we have developed a con- formational analysis technique which combines simulated annealing (SA) (Kirkpatrick, et al., Science, 220:671 (1983); Simulated Annealing and Optimization, M.W. Johnson, Ed., American Sciences Press, Syracuse, N.Y.  
35       (1988)); Wilson et al., Tetrahedron Letters, 4343 (1988); Wilson et al., Proceedings of the Seventh

Workshop on vitamin D, Walter de Gruyter Co. (1988); Wilson and Cui, Biopolymers, 29:225 (1990); Wilson et al., J. Comp. Chem., 12, 3, 342 (1991)) and biased sampling (Kalos and Whitlock, Monte Carlo Methods vol 1, Wiley, New York, N.Y. 1986) into a type of learning algorithm (Caudill, Expert, 12/89, 4/90, 6/90; Judd, in Neural Network Design and the Complexity of Learning, MIT Press, Cambridge, MA (1990); Statistical Mechanics of Neural Networks, Luis Garrido, Ed, Springer-Verlag New York, N.Y. (1990); Lacey, Ed., Neural Networks, Tetrahedron Computer Methodology, 1990, 3). The method is a 2-stage process made up of a learning phase and an implementation phase. The learning phase starts by randomly sampling the dihedral space of all flexible bonds using the simulated annealing algorithm (Kirkpatrick et al., Science, 220:671 (1983); Simulated Annealing and Optimization, M.W. Johnson, Ed., (American Sciences Press, Syracuse, N.Y. (1988)). The entire 360° continuous dihedral space of all flexible torsional angles is sampled in accordance with the fundamental hypothesis of equal a priori probabilities (Tolman, in The Principles of Statistical Mechanics, (Dover Press, New York (1971)). To provide our knowledge-base, multiple SA runs are performed and for each step, the chosen dihedral, value of the chosen dihedral and conformation energy at that step are recorded (F. Guarnieri, Ph.D., Thesis, New York University (1992)). This series of log files is converted into population distributions by summing and/or averaging the number of hits in ten degree intervals. The conformation space of the antiinflammatory agent LTB<sub>4</sub>, which has fourteen rotatable dihedral angles, gives us the Flex-Map (Wilson and Guarnieri, Tetrahedron Lett., 32:3601 (1991)) plots of these fourteen bonds. One typical Flex-Map is shown in Figure 2. These plots contain information of the overall

population distribution for each rotatable bond as a function of temperature. Since the whole molecule is in flux with all energetic interactions taken into accord at every step, the Flex-Maps are mean field population distributions with no approximations. Hence, they are a true canonical ensemble with respect to all flexible torsional states of the molecule. These maps rapidly reveal occupied regions of dihedral space and "dead zones" which are totally devoid of conformations at any temperature. These "dead zones" are the key to why it is so difficult to search the conformation space of flexible molecules with many rotatable torsionals. Most methods sample from the whole space throughout a conformation search. Clearly these dihedral distributions, which we now call conformational memories, indicate that sampling from many regions is a complete waste of time.

It is self-evident that it will be vastly more efficient to sample from the smaller space obtained from the elimination of "dead zones" compared to sampling the original space. The key point is to make sure that thermally accessible regions are not erroneously labeled as "dead zones" because the second phase of the simulation would be flawed. Thus, in the first part of the simulation, care must be taken to insure a good sampling. This is why repeated simulated annealing runs are performed in the initial phase. Multiple runs from different random starting geometries using high temperatures and large Monte Carlo steps have the best chance of sampling in every region. We would like to point out that a comparable simulated annealing strategy was shown to be capable of searching the entire conformation space of cycloheptadecane (Wilson and Guarnieri, Tetrahedron Lett., 32:3601 (1991); Guarnieri and Wilson, Tetrahedron, 48:4271 (1992); Guarnieri et al., J. Chem Soc., Chem. Comm.,

21:1542 (1991)) in less than 48 hrs on a microvax. In contrast, the aforementioned effort (Saunders et al., J. Amer. Chem. Soc., 112:1419 (1990)) using most known search methods took about 2 CPU years on a microvax. For more complicated systems such as LTB<sub>4</sub>, we are confident that compilations of repeated runs started from different configurations, using different random number seeds, and initialized with a thermal energy of over 1000K, reveals the populated and unpopulated regions of the 14 dimensional torsional space of LTB<sub>4</sub>. In fact, the unpopulated regions are revealed particularly early on in the simulation. In compiling 5, 10, 15 and 20 runs, the ratios of the populated regions change (to a very small degree in going from 15 to 20 runs), but there is virtually no change in unpopulated regions throughout this progression. For LTB<sub>4</sub> the unpopulated regions make up more than half of the total conformational space at 200K. A sampling strategy which avoids these "dead zones" would reduce the volume of conformation space that needs to be searched from  $360^{14}$  to less than  $180^{14}$ , where 14 is the dimensionality of the space, and 360 is the extent of one dimension.

Smart-Simulated Annealing: The Implementation Phase

The implementation phase involves utilizing the information contained in the 14 conformational memories, an example of which is shown in figure 2. This is done by again running the SA Metropolis algorithm, but instead of selecting new trial conformations at random over the whole dihedral circle, we select new trial conformations by sampling only from populated regions of the conformational memory for each bond at a given temperature. To search for low energy conformations, the populations at 200K were chosen. We call this technique of biased sampling with simulated annealing Smart-SA.

-11-

A new procedure is needed to sample a dihedral space embedded with "dead zones." In order to carry out this biased sampling, it is necessary to map the uniformly distributed random numbers produced by standard random number generators onto the conformational memories. This process is illustrated in figure 3. The conformational memory in figure 2 may be approximately by a classic three state model. Figure 3 shows mapping of random numbers into this distribution. To perform this mapping, our algorithm requires information on the number states, the interval and population of each state. (In the example, the number of states is actually four instead of three because dihedral space goes from -180 to +180). The shaded regions are the "dead zones", and thus are never sampled. Since the first region has a 1/3 probability of being surveyed, if the generated random number is between zero and 1/3, the new dihedral is selected from this first region. The exact value that this bond will be set to is obtained by starting at the point on the ordinate at the value of random number, moving horizontally until the dotted line is met, dropping a vertical from that point, and selecting the dihedral value that arises from the intersection of the vertical with the abscissa. For example, a probability of 0.60 maps to a dihedral value of 65°. This angle is passed to the dihedral driver to construct the new conformation. Once this conformation is created, the algorithm passes back to our standard simulated annealing routines (Kirkpatrick et al., Science, 220:671 (1983); Simulated Annealing and Optimization, M.W. Johnson, Ed., American Sciences Press, Syracuse, N.Y. (1988); Wilson et al., Tetrahedron Letters, 4343 (1988); Wilson et al., Proceedings of the Seventh Workshop on Vitamin D, Walter de Gruyter Co., (1988); Wilson and Cui, Biopoly-

mers, 29:225 (1990); Wilson et al., J. Comp. Chem. 12, 3, 342 (1991).

### Results

5       The LTB<sub>4</sub> problem above was run using the same SA control data as previously reported (Kirkpatrick et al., Science, 220:671 (1983); Simulated Annealing and Optimization, M.W. Johnson, Ed., (American Sciences Press, Syracuse, N.Y. (1988)). Ten runs of SA and  
10   Smart-SA were carried out on LTB<sub>4</sub> (500 steps at 25 temperatures). The convergence results are shown in figure 4. Much faster lowering of the conformer energy occurs producing rapid convergence.

      On a system as complicated as LTB<sub>4</sub> it is impos-  
15   sible to prove that a conformation search is complete. One usual measure of comprehensiveness is to perform repeated searches using different initial conditions until the output of several searches produce the same results. Ten ordinary simulated annealing runs on LTB<sub>4</sub>  
20   produced ten different low energy conformations. Ten runs of Smart-SA with the conformational memories taken at 200K produced two conformations (6 of one and 4 of the other) which were both lower in energy than any of the ten conformational found by ordinary simulated  
25   annealing.

### Computational Details

      Since many of the computational details have been reported, (Wilson et al., Tetrahedron Letters, 4343 (1988); Wilson et al., Proceedings of the Seventh  
30   Workshop on Vitamin D, Walter de Gruyter Co., (1988); Wilson and Cui, Biopolymers, 29:225 (1990); Wilson et al., J. Comp. Chem. 12, 3, 342 (1991); F. Guarnieri, Ph.D. Thesis, New York University, 1992; Wilson and Guarnieri, Tetrahedron Lett. 32:3601 (1991)) only a  
35   brief summary will be outlined. All runs were started with beta=0.11. This corresponds to a temperature of

1093K given that  $\beta = 1/(RT)$  where  $R = 8.314 \text{e3 kJ/mol}$ . After every block of steps  $\beta$  is multiplied by 1.1 to reduce the temperature. In theory, the cooling schedule should be controlled and varied as a function of the heat capacity. In practice, we have found that balancing the need for slow cooling and obtaining acceptable CPU performance is a reasonable compromise. At each step the dihedral angles that are rotated to create the trial structure are noted. Whether the trial configuration is accepted or rejected, the identity of the rotated dihedrals, the extent of the rotation, the value of the energy of the trial configuration and the new dihedral values if the trial conformation is accepted or the old dihedral values if the trial conformation is rejected are recorded to the log file in temperature blocks. One log file is created for each run. A utility program inputs all of this raw data and combines it according to temperature blocks. This data is output in comma delimited format so that it can be imported into deltagraph (Deltagraph TM version 1.0, Copyright Deltapoint, Inc., 200 Heritage Harbor, Suite G, Monterey, CA 93940) which is used to plot the conformational memories. Using another utility program or manually, a temperature slice from the conformational memory is extracted. In the second phase of the simulation the sampling is done from a subroutine that performs the calculation shown in figure 3 instead of just using the standard random number generator.

At the outset of the study we were faced with the nearly intractable 14-dimensional problem. The learning phase of the simulation reveals that about 60% of the entire conformational space is unpopulated "dead zones" at 200K. Going into the implementation phase of the simulation, we were able to reduce the volume of the conformation space that needed to be sampled by

many orders of magnitude. Additionally, since several  
dihedrals remained exclusively in a trans conformation  
at all temperatures throughout the many learning phase  
simulations, we were also able to reduce the dimension-  
5 ality of the search in the implementation phase by  
setting these dihedrals to a constant value of 180 with  
no loss of generality. To reiterate, Smart-SA  
(simulated annealing with biased sampling) allows a  
considerable reduction in the conformational space  
10 which needs to be sampled. Hence, much larger systems,  
which have been generally considered computationally  
intractable, may now be studied.

Having used this technique with much success in  
the area of conformation searching, we have begun  
15 exploring Smart-SA applications to quantitative  
problems such as free energy simulations. Preliminary  
results with sampling proportional to the height of the  
populated peaks has proven dependent upon the length of  
the learning phase. Only after getting quantitative  
20 convergence in the ratio of the populations are the  
final results invariant. This problem can be traced to  
a violation of detailed balance. Preliminary results  
of simulations sampling equally from all populated  
regions (with no sampling from "dead zones"), which  
25 should not violate detailed balance, have proven suc-  
cessful.

#### Example: Identification of Active GnRH Analogs

The key physiological role of the gonadotropin-  
30 releasing hormone ([pGlu1-His2-Trp3-Ser4-Tyr5-Gly6-  
Leu7-Arg8-Pro9-Gly10-NH<sub>2</sub>];GnRH) as a mediator of  
neuroendocrine regulation in the mammalian reproductive  
system has made it the object of intense study for  
several decades. The ability of GnRH and its analogs  
35 to modulate the pituitary-gonadal axis has made them  
essential therapeutic agents in the treatment of a



variety of disorders ranging from infertility to prostatic carcinoma (Casper, Can Med Assoc J. 144:153-160 (1991); Barbieri, Trends Endocrinol Metab 3:30-34). Conformational studies have played a central role in the quest for understanding the structural basis for the activities of these peptides, as well as in attempts to design new analogs with improved pharmacological properties. Investigations of the biological mechanisms underlying the actions of the GnRH are quite difficult because small peptides are extremely flexible. Spectroscopic techniques, for example, indicate that a multitude of interconverting conformers exist simultaneously. In an attempt to pare away some of the many thermally accessible but biologically irrelevant conformations, several investigators have synthesized restricted GnRH analogs (Rizo et al., J. Amer. Chem Soc., 114:2852 (1992); Bienstock, et al., J. Med Chem., 36:3265 (1993)). This approach has proven useful for defining some structural motifs of antagonists of the hormone. Obtaining comprehensive detailed molecular conformational properties, however, such as the specific dihedral values of the bioactive forms, is a formidable task given that the GnRH has 35 rotatable bonds as shown in figure 5.

The inherent complexities of small flexible peptides has motivated studies which combine computational and experimental techniques (Young and Hicks, Biopoly., 34:611 (1994)). The computational method of choice is molecular dynamics. While dynamical techniques are capable of revealing short time scale molecular motions, these methods are generally incapable of exploring the ensemble of conformational states that exist in flexible molecules (Guarnieri and Still, J. Comp. Chem., 15:1302 (1994)). To explore the whole ensemble of conformational states that exist in the GnRH, we have used the recently developed technique of

conformational memories (Wilson and Guarnieri, Tetra-  
hedron Lett., 32:3601 (1991)). Here we show that  
application of this technique can yield converged  
dihedral populations of all 35 rotatable bonds of the  
peptide. GnRH with no approximations. Samples from  
the conformational memories using the conformational  
memory biased sampling technique were used to charac-  
terize the conformational families of GnRH and several  
of its analogs, in an aqueous environment modeled with  
the generalized Born/surface area (GB/SA) method.  
(Still et al., J. Amer. Chem Soc., 112:6127 (1990)).  
This analysis reveals the conformational preferences of  
the GnRH and its analogs, and suggests some of the  
structural determinants for their biological function.

15

#### Method of Conformational Memories

The simulation technique of conformational  
memories is a two stage process consisting of an  
exploratory phase and a biased sampling phase. In the  
exploratory phase repeated runs of Monte Carlo simu-  
lated annealing (MC/SA) (Kirkpatrick et al., Science,  
220:671 (1983)) are carried out in order to map out the  
entire conformational space of the flexible molecule.  
The construction of Conformational memories described  
below has been interfaced with the Macromodel (Mohamadi  
et al., J. Comput. Chem. 11:440 (1990)) molecular  
modeling package version 5.0 so that the continuous  
GB/SA solvent model (Still et al., J. Amer. Chem. Soc.  
112:6127-6129 (1990)), and the recently developed amino  
acid backbone torsional potentials (McDonald et al.,  
Tetra. Lett. 33:7743-7746 (1992) from the Macromodel  
package could be used in the present conformational  
study. As applied here, the MC/SA protocol for the  
exploratory phase was designed with a starting tempera-  
ture of 2070 and a cooling schedule of  $T_{n+1}=0.9*T_n$  for  
nineteen discrete temperature points. At each tempera-

ture 10,000 steps were applied to the 35 rotatable bonds (fig. 5), cooling the system to a final temperature of 310K. Trial conformations in the MC/SA routine were generated by randomly picking 2 rotatable bonds from among the 35, rotating each bond by a random value between +/-180 degrees, and accepting or rejecting the trial conformation according to the standard Metropolis (Metropolis et al., J. Chem Phys., 21:187 (1953)) criteria with a Boltzmann probability function defined at the given temperature. After each step, whether the conformation was accepted or rejected, the data for the rotated bonds, the extent of rotation, the energy, and the value of the dihedral angles are recorded to a "log file". An example of the output to the log file is given in Table 1. In this example, the first group of entries, corresponding to the first two lines, is the result of a rejected step as indicated by the zeros in the first column. The second and third columns identify the atom numbers of the bonds that were rotated to create the trial move (in this example atom40-atom41 and atom47-atom48). The fourth column lists the extent of rotation of the torsion angle in degrees. The fifth column lists the total energy of the structure. The sixth column holds the current dihedral value of the bond.

The second group of entries in Table 1, corresponding to the next two lines, lists the results of a trial rotation that was accepted as a new conformation (as indicated by the digit one in the first column). The current dihedral values given in the last column are the new values of the newly accepted conformation.

Each run of MC/SA consists of a random walk of 190,000 steps (19 temperatures, 10,000 steps per temperature). Because two lines of data are added to the log file for each Monte Carlo step, a single run creates a file of 380,000 lines. To explore the

conformations of the GnRH peptide in GB/S A water, we performed 157 of these simulations, creating log files of the different random walks. A 157 run MC/SA simulation requires about 12 days of computation on an SGI Challenge 200 MH<sub>3</sub> workstation.

To obtain structural information from this large amount of data the log files are used as input to a program (called Flex) that sorts, merges, and compacts the data in several ways. Since the simulations were done at 19 temperatures for each peptide, application of Flex first sorts and merges the data from all log files into 19 temperature blocks. Subsequently, within each temperature block, the data are partitioned into 35 bond blocks, one for each rotatable bond. For each rotatable bond, the dihedral angle space is partitioned into 36 ten degree intervals. From each line of data for a given bond at a given temperature, the program records the number of times that the bond dihedral angle value belongs to one of the ten degree buckets, i.e. a "Conformational Memory". Finally, the Flex program produces a 19x36 (recording 19 temperatures by 36 10-degree diheral intervals with normalized populations) spread sheet for each of the 35 rotatable bonds of the GnRH peptide. An excerpt of one of these spread sheets is given in Table 2. The spreadsheets are imported into Delagraph (TM Version 1.0, Copyright Deltapoint, Inc. 200 Heritage Harbor, Suite G, Monterey, CA 93940, (1987)) for plotting and graphical representation of the data in the spreadsheets are given in Figures 6 A-D. Across the top of the spread-sheet are the dihedral angle values from -170° to 180° which label the y-axes of Figures 6 A-D (note that the spreadsheet fragment is cut off at -100). In the first column are the 19 temperatures which range from 2070 to 310 which label the x-axes in Figures 6 A-D. The value in a spread sheet position corresponding to a given

temperature with a given ten degree dihedral bucket is the population percentage which is plotted on the z-axes of Figures 6 A-D.

The procedure for creating conformational memories for the dihedral angles results in an enormous compression of the large volume of data needed to describe a 35 dimensional hypertorsional space. The condensation of the information in Deltagraph plots yields identifiable structural motifs. For example, bond 4 shown in figure 6a, has a classic three state distribution: trans, gauche+ and gauche-; bond 6, the phi angle of residue 3 (Fig. 6b), has a continuous population distribution over a very large range from about -60 to -180 degrees and no population in the other regions at any temperature. In contrast, bond 7, the psi angle of residue 3 (Fig. 6c), has a narrow all trans distribution. The distribution of bond 35 (Fig. 6d), favors a trans conformation, but maintains significant population over the entire dihedral circle at all temperatures. The construction of conformational memories has been interfaced with the Macromodel molecular modeling package (Mohamadi et al., J. Comp. Chem., 11:440 (1990)) version 5.0 so that the continuum GB/SA solvent models and the recently developed amino acid backbone torsional potentials (McDonald and Still, Tetra. Lett., 33:7743 (1992)) from the Macromodel package could be used in the present conformational study.

### 30 Convergence of Conformational Memories

By including the results from multiple explorations of all possible combinations of dihedral angle values for all rotatable bonds of the molecule, the thirty-five conformational memories provide a complete mapping of the conformational space of GnRH with no approximations, as long as the calculated popula-

tions are converged. In the original formulation of the method, population convergence was identified as the difficult and crucial aspect of forming conformational memories (Wilson and Guarnieri, Tetrahedron Lett., 32:3601 (1991)). Because the second phase of the simulation, the biased sampling, explores only the parts of the conformational space identified as populated regions, Population convergence ensures that regions that could be thermally accessible are not erroneously labeled as being unpopulated. The correct identification of the populated regions is essential for the second phase of the simulation, because the biased sampling only explores populated regions of the conformational space.

Population convergence for the GnRH was confirmed in three different ways: by creating conformational memory difference maps for simulations of different length, by analyzing intrinsic symmetry; and by showing that there is no significant difference in the populations of actual structures of GnRH created from Conformational Memories obtained from 25, 50, 75, 100 and 157 independent MC/SA runs. Figure 3 shows the Conformational Memory difference maps for dihedral angle 1, comparing simulation lengths of 10, 25, 50, 75 and 100 runs. The difference map in Figure 3a is created by subtracting the Conformational Memory obtained from a 25 run MC/SA simulation from a 10 run MC/SA simulation, in Figure 3b the difference is between 50 and 25 runs, Figure 3c shows the difference between 75 and 50 runs, and Figure 3d is the difference map between 100 and 75 runs. The progression clearly shows the convergence. The other dihedral angles have very similar difference maps for this sequence of comparisons.

A second measure of convergence is symmetry. Because dihedral angle 17 has a 2-fold axis of sym-

metry, it is expected that the dihedral space of this bond will have symmetric population distributions centered at -90 and 90 degrees. A temperature slice at 310K of this dihedral for 25, 50, 75 and 100 run MC/SA simulations is shown in Figure 8. The population distributions clearly conform to the symmetry considerations.

The third indication of convergence is the finding (see below) that biased sampling from Conformational Memories created from 25, 50, 75, 100 and 157 MC/SA runs yield very similar profiles of GnRH.

#### Biased Sampling From Conformational Memories: Elimination of Barriers

15

Once the conformational memories are established, a new Monte Carlo search is performed at 309K, sampling only from the populated regions. Because about 50% of the torsional space of the 35 bonds is populated at 310K, so that the conformation space that needs to be explored in the biased sampling phase of the simulation has been reduced without approximations, by many orders of magnitude. Table 3 is an excerpt of the probability matrix for GnRH at 310K. The dimension of this probability matrix is 35x36 for the 35 rotatable bonds partitioned into 36 buckets over the 360 degree dihedral space (note that only 11 of the 36 dihedral buckets and only 16 of the 35 rotatable torsional angles are shown in Table 3). The first line indicates that at 310K bond 1 is found in the -180 to -170 dihedral interval 10.1% of the time. In contrast, the seventh column of the first row indicates that bond 1 is never found in the dihedral interval -120 to -110 at 310K.

35

The two stage process of developing Conformational Memories and then performing the biased sampling from

these distributions is necessary in order to sample the entire conformation space of the molecule. An obviously simpler alternative would be to limit the conformational exploration to standard Metropolis Monte Carlo at 310K and monitor the development of the random walk over torsional space. However, this simulation constitutes the last step in the development of the Conformational Memories for the temperature of 310K; it is clearly inadequate, as indicated by the acceptance rate. The acceptance rate is about 28% at 2070K, with a step size chosen randomly within the interval of +/- 180 degrees and rotating two dihedrals selected randomly at each step. At 310K, using the same parameters, the acceptance rate falls below 2%. Therefore, the sampling of the 35 dimensional dihedral space would be incomplete if these parameters were used for the Monte Carol random walk procedure at 310K. Even if the random interval from which trial configurations are sampled were reduced to +/-30 degrees (to increase the acceptance rate), sampling would still be insufficient because the majority of new conformations would be in the local area of the previous conformation. The +/- 180 degree step size was deliberately chosen so that new conformations can be created by jumping between wells without having to climb over barriers. A single simulated annealing run cannot be expected to cover such a vast space, but cumulations of multiple runs while each of the runs performs a different random walk can be shown to converge, as illustrated in Figures 7 and Figure 8.

The restriction of the sampling to the populated regions identified in the previous step (i.e., the Conformational Memories) is achieved by partitioning the 0-1 interval of the random number generator into the 36 parts which correspond to the 36 separate 10-degree intervals for each rotatable dihedral angle.



The partitioning of the random number generator is proportional to the population of the 10-degree bucket. New biased trial conformations are generated by randomly choosing two rotatable bonds, generating a new random number for each bond, determining to which of the 36 intervals each new random number for each bond belongs, and driving the dihedrals to the appropriate intervals. The exact value of the new dihedral is determined by a linear interpolation. This procedure is illustrated in Figure 9.

A major advantage of the Conformational Memory biased sampling method is that partitioning the random number generator among the populated intervals results in a sampling technique that eliminates the barrier-crossing problem. During the biased sampling random walk, a new trial configuration is sampled from the Conformational Memory, which can be any part of the populated dihedral space, and then the trial conformation is created by driving the current structure to the appropriate configuration. Hence, the notion of a barrier restricting access to any part of the conformational space is eliminated in this procedure. Because Conformational Memories are mean field population distributions, the correlations among the different flexible torsional angles have been submerged in the averaging process. Nevertheless, the Conformational Memory biased sampling technique does preferentially bring together the higher probability regions of the different dihedrals. Thus, the method introduces average correlations among the different dihedral angles during the selection process, while accessing all populated regions. It is important to note that the original formulation of the Conformational Memories biased sampling technique (Guarnieri and Wilson, J. Comput. Chem 16:648-653 (1995)) violates detailed balance. Here, we have corrected the biased sampling

so that it obeys detailed balance by multiplying the Boltzmann function used in the Metropolis test with the factor  $P_{1old} \cdot P_{2old} / (P_{1new} \cdot P_{2new})$ , where  $P_{1old}$  and  $P_{2old}$  are the population percentages of the ten degree intervals of the Conformational Memories of the two dihedral angles in the current conformation of the random walk (because in this example two dihedrals per step are changed).  $P_{1new}$  and  $P_{2new}$  are the corresponding population percentages of the new dihedral values for these angles in the new trial conformation.

#### Development of Conformational Families

We performed several sequences of biased sampling runs at 310K to determine the best and simplest way to create representative conformational families for the GnRH peptide. The first run was a 10,000 step MC random walk using the Conformational memory biased sampling technique with uniform sampling of 100 structures (1 sample every 100 steps). The second run was a 50,000 step MC random walk using the Conformational Memory biased sampling technique with uniform sampling of 100 structures (1 sample every 500 steps). The third and fourth runs were 100,000 and 500,000 step biased sampling runs also sampling 100 structures in the same manner. Each batch of 100 structures was analyzed with the program XCluster (Shenkin and McDonald, J. Comput. Chem 15:899-916 (1994)). XCluster inputs the series of 100 conformations and computes the RMS difference between all possible pairs of conformations. Structures 2-100 of the input sequence are then reordered based on increasing RMS deviation. In the new ordering, considering all 100 conformations, conformer 2 has the smallest RMS deviation from conformer 1, and conformer 3 has the smallest RMS deviation from conformer 2, etc. Xcluster then produces a graphical representation of the RMS deviations between every pair

of conformers. Since the conformations have been rearranged so that the RMS deviation between nearest neighbors is minimized, any large jump in RMS deviation between nearest neighbors is indicative of a large structural change and hence identifies a new conformational family. As described below, we settled on 500,000 steps for the subsequent biased sampling runs. We then performed these biased sampling runs using Conformational Memories created from 25, 50, 75, 100 and 157 run MC/SA simulations.

### III. Results And Discussion

#### Conformational Families of GnRH

The 500,000 step biased sampling runs for GnRH with a sampling rate of 1 every 5,000 structures require 4.3 hours per run on a 200 MHz SGI Challenge workstation. Structures from the 500,000 step biased sampling run were clustered in conformational families as described above. A backbone trace of representatives from the 5 families with very distinct backbone conformations that emerged from this procedure is shown in Figure 10. Notably, similar results were obtained regardless of the origin of the Conformational Memories from MC/SA simulations of 25, 50, 75, 100 or 157 runs. Families of conformations having a beta-turn between residues 5-8 occur with a frequency of approximately 70%. A distribution showing a superimposition of 70 of these structures is illustrated in Figure 11 (GnRH is colored in red, with Arg8 colored in green). The beta-type turn common to all the structures in this family is clearly evident (Fig 11). In contrast, families which have an extended backbone, occur with a frequency of about 5%. The distribution of side chain orientations of Arg8 in all conformational families is wider than that of any other residues in GnRH. The results of Struthers et al, (Proteins: Structure, Function and

Genetics 8:295-304 (1990)) from the examination of different GnRH analogs seem to indicate that an arginine is required as part of the pharmacophore. The present results, on the other hand, may indicate that the role of Arg8 in the receptor interaction of GnRH could relate to the backbone conformation, rather than to its participation in a recognition pharmacophore.

It is noteworthy that biased sampling runs of 10,000-25,000 steps resulted in large (unconverged) fluctuations in the ratio of beta-turn to extended backbone conformations. However, more extended biased sampling runs of 100,000-500,000 resulted in negligible fluctuations in the ratio of beta-turn to extended backbone conformations. Although it appeared from our calibration studies that 100,000 step biased sampling runs are sufficient, we chose to carry out the more extensive 500,000 step biased sampling runs for all the calculations presented here.

#### 20 Conformational Families of Lys8-GnRH

The Lys8 analog of GnRH had been constructed to explore the role of Arg8 in molecular recognition of GnRH by its receptor (Karten and Rivier, Endo. Rev. 7:44-66 (1986); Millar et al., J. Biolog. Chem. 264:21007-21013 (1989)). Mutation studies of GnRH receptors from various species have implicated Arg8 as being important for mammalian hormone-receptor recognition (Flanagan et al., J. Biolog. Chem. 269:22636-22641 (1994)). To analyze the structural implications of Arg8 for the activity of GnRH, we compared the conformational profile of the peptide hormone with that of the mutant Lys8-GnRH which is known to be a low affinity GnRH agonist. In contrast to the wild type hormone, the major conformational family of the Lys8-GnRH congener was found to have an extended backbone, while the beta-turn conformation exists as a very minor

family. A backbone trace of a representative of each family is shown in Figure 12. The family of conformations represented in Figure 12a has an extended backbone and occurs with a frequency of greater than 70%. The Lys8-GnRH family that has a beta-type turn conformation of the backbone (Figure 12b) which is virtually identical to the major conformational family of the GnRH (Figure 10a), has a probability of only about 3%. A distribution of the members of the predominant Lys8-GnRH family superimposed upon each other is shown in Figure 13, with the entire molecule shown in red, except for Lys8 which is colored green. Because the Lys8-GnRH has a low affinity for the GnRH receptor, but elicits the same response once it interacts with the receptor, it is tempting to suggest that adoption of a large population of beta-type turn conformation is a key requirement for hormone-receptor recognition. This inference agrees with earlier proposals in the literature, and is supported by results from additional Conformational Memories simulations on the structural characterization of eight other GnRH analogs that exhibit different distributions between the beta-turn like structures and the fully extended conformations of the backbone (Guarnieri et al., unpublished results). It is particularly noteworthy that our simulations lead to the same conclusions regarding the importance of the bent structure that were drawn from their combined NMR and molecular dynamics studies of conformationally constrained GnRH analogs (Struthers et al., Proteins: Structure, Function and Genetics 8:295-304 (1990)).

#### Structural Comparison To a Constrained GnRH Analog

To test the key inference from the present simulations of GnRH analogs, regarding the correlation between the population of beta-type turn structure and

affinity for receptor, we compared several samples from the most populated conformational family of GnRH obtained from Conformational Memories to a structurally constrained cyclic decapeptide GnRH analog (Baniak et al., Biochem 26:2642-2656 (1987)). The conformation of this cyclic decapeptide was determined from NOE data using 2D NMR techniques (Baniak et al., Biochem 26:2642-2656 (1987)). These experimental studies concluded that residues 6 and 7 formed a type II beta-turn and residues 1 and 2 formed a type II beta-turn. Additionally, it was concluded that a weak hydrogen bond existed between the Arg8 -NH and the Tyr5 -CO, and a stronger hydrogen bond between the D-Trp3-NH and the beta-Ala10 -CO. To allow for the comparison, a structure of this GnRH analog was built in Macromodel 4.5 according to the specifications (Baniak et al., Biochem 26:2642-2656 (1987)), and using the beta-turn definitions of Hutchinson and Thornton (Hutchinson and Thornton, Protein Science 3:2207-2216 (1994)). This reconstructed GnRH analog was compared to the GnRH structures obtained from the Conformational Memories described above.

Several of the members of the major conformational family of the GnRH obtained from the Conformational Memories were selected at random and superimposed on the reconstructed geometry of the analog, using the 11 backbone atoms from the Tyr5 -CO to the -N of Pro9. All computationally derived structures superimposed on the reconstructed structure with RMS deviation in a range of 0.6-0.8 Å. An illustration of the superimposition is shown in Figure 14. Clearly, the computationally derived structure is closely related to the reconstructed backbone of residues 5-8 of the experimentally derived peptide structure. The structures diverge between the N-terminus and residue 4,

and a superimposition of all backbone atoms results in a 5 Å RMS deviation.

#### GnRH Conformations From a Buildup Procedure

5       Recently, (Nikiforovich and Marshall, Int. J. Peptide Protein Res. 42:171-180 (1993)) constructed low energy conformations of GnRH using the ECEPP program (Dunfield et al., J. Phys. Chem. 82:2609-2616 (1978)). We have reconstructed eight conformations from the published list of backbone dihedral angles and a list of side chain dihedrals graciously provided by the authors (Nikiforovich and Marshall, Int. J. Peptide Protein Res. 42:171-180 (1993)). The energies of these reconstructed peptide structures were compared with  
15       representatives from the three major families of GnRH found using Conformational Memories. The optimal geometries of GnRH obtained from the two computational methods were quite different, and the energies of the eight conformations calculated from ECEPP were 300-400  
20       kJ/mol (20-25%) higher than those calculated from the conformations generated using Conformational Memories. It is unlikely that this large difference can be attributed solely to the use of different force fields in the definition of optimal conformations, since a recent  
25       comparative study resulted in very similar low energy Met-Enkephalin structures (Montcalm et al., J. of Mol. Str. (Theochem) 308:37-51 (1994)). However, a major source of difference may be the use of a GB/SA water model in the Conformational Memories approach, and  
30       perhaps a more complete exploration of the conformational space.

#### Exploration of the Unpopulated Regions

35       As a stringent test of the completeness of the conformational exploration, we performed extensive sampling from the unpopulated regions of the

Conformational Memories for several key dihedral angles involved in the formation of the beta-turn of the GnRH. With one exception, this sampling produced high energy structures in all cases, as expected. The one  
5 interesting exception occurred during the sampling of the unpopulated regions of the phi angle of Gly6. This sampling produced a structure only 20 kJ/mol higher in energy than the best GnRH structure. The dihedral value came from a bin that had a 0.6% population at  
10 345K, but had a 0% population at 310K and therefore was not included in the populated portion from which the MC biased sampling was done. A simple way to avoid missing a very low probability low energy structure when performing the biased sampling at 310K, is to use  
15 the probability weights from a higher temperature. Our exploration of the unpopulated regions of the phi angle of Gly6 at temperatures 100-200K above 310K eliminated this problem. The small drawback, however, is that 44% of the dihedral space of Gly6 is unpopulated at 310K  
20 and only 33% is unpopulated at 473K. Thus, a safety factor during the biased sampling run involves exploring about 10% more dihedral space per rotatable torsion, but ensures the enclosure of all populated areas. Conformational regions that exhibit 0%  
25 population in the calculation of the isolated peptide in water at 310K may still be of biological importance, if some of these conformations can be induced by the interaction energies of the peptide with the receptor. The finding that regions unpopulated at 310K are in  
30 fact populated at temperatures higher by only 100K (corresponding to an energy difference of only a fraction of a Kcal/mol), indicates the feasibility of such "receptor-induced" conformations.

### 35 Conclusions



Applied to the decapeptide hormone GnRH, the method of Conformational Memories was shown to provide a powerful practical solution to the complex problem presented by the flexibility of polypeptides with a large number of conformational degrees of freedom. With the study of the flexible decapeptide, the method was shown to be capable of achieving complete sampling of the conformational space, to converge in a very practical number of steps, and to be capable of overcoming energy barriers efficiently.

The results of the conformational study support a relation between the beta-turn structure identified as the major conformational family of GnRH, and high affinity for the GnRH receptor. While these inferences were inherent in the results for earlier investigations of conformationally restricted GnRH analogs, the present study provides unbiased support for this mechanistic hypothesis based on a complete exploration of the conformational space of the peptide hormone itself and its unconstrained congeners. Because the method seems to have produced the lowest energy conformers reported for GnRH from a full exploration that is economical and practical, its general application to the study of peptide structure-function relations should continue to produce important mechanistic insights and powerful guides for ligand design.

#### TABLE 1

A sample of the output collected in the history files of the simulated annealing random walks. Column 1 indicates if the data is produced from an accepted or rejected step with 0=rejected and 1=accepted. The second column lists the pair of atom number identifying the dihedral angles that were rotated to produce the trial structure. The third column lists the extent to which the dihedral was rotated in order to create the trial structure. The fourth column lists the energy of

the current conformation (the energy of the original structure if rejected or the new structure if accepted). The fifth column lists the current dihedral values of the conformation (the dihedral angle of the original structure if rejected or the new structure if accepted).

**Table 2**

A sample of a conformational memory spreadsheet. The first row labels the dihedral circle across the y-axis. The first column labels the temperatures across the x-axis. Each cell contains the population corresponding to a given temperature and a given 10 degree dihedral bucket which is plotted on the z-axis. Note that the columns of the spreadsheet are cut off after -

40 degrees.

**Table 3**

Excerpt from the population probability matrix for GnRH at 310K. The dimensions of this matrix are 35x36 (35 rotatable dihedral angles, with the population distribution of each angle broken into 36 intervals of 10 degrees). Note that only 14 rows and 11 columns of this matrix are shown in the Table.

Various publications are cited herein, the contents of which are hereby incorporated by reference in their entireties.

**Table 1. Sample of a conformational memory history file.**

| 1=Acc<br>0=Rej | Rotated Dihed<br>Atom1-Atom2 | Extent of<br>Rotation | Energy  | Dihed Values<br>Conformation |
|----------------|------------------------------|-----------------------|---------|------------------------------|
| 0              | 40 41                        | 63.86                 | -276.38 | 176.53                       |
| 0              | 47 48                        | -28.67                | -276.38 | -77.32                       |
| 1              | 90 91                        | 3.48                  | -252.98 | 168.15                       |
| 1              | 3 4                          | 11.96                 | -252.98 | -77.33                       |

**Table 2. Excerpt of a Conformational Memory Spreadsheet**

| Temp  | -180° | -170° | -160° | -150° | -140° | -130° | -120° | 110° | -100° |
|-------|-------|-------|-------|-------|-------|-------|-------|------|-------|
| 1500K | 4.0   | 3.7   | 3.2   | 2.5   | 2.1   | 1.5   | 1.3   | 1.3  | 1.1   |
| 1350K | 4.7   | 3.7   | 3.7   | 2.4   | 2.3   | 1.5   | 1.4   | 1.1  | 0.9   |
| 1215K | 4.0   | 4.0   | 3.7   | 2.4   | 2.0   | 1.1   | 1.0   | 0.9  | 0.9   |
| 1094K | 4.5   | 3.6   | 3.6   | 2.4   | 1.9   | 1.3   | 0.9   | 0.9  | 1.0   |
| 984K  | 5.4   | 4.6   | 3.9   | 2.2   | 1.7   | 1.1   | 0.8   | 0.9  | 1.0   |
| 886K  | 5.3   | 4.4   | 3.4   | 2.6   | 1.9   | 1.2   | 0.9   | 0.8  | 0.5   |
| 797K  | 5.8   | 4.9   | 3.3   | 1.9   | 1.2   | 0.9   | 0.6   | 0.4  | 0.5   |
| 717K  | 6.0   | 4.2   | 3.5   | 2.4   | 1.4   | 0.9   | 0.5   | 0.2  | 0.4   |
| 646K  | 5.6   | 4.6   | 3.5   | 2.2   | 1.0   | 0.7   | 0.3   | 0.3  | 0.4   |
| 581K  | 6.3   | 5.1   | 2.9   | 1.9   | 1.4   | 0.4   | 0.4   | 0.2  | 0.2   |
| 523K  | 6.1   | 3.9   | 3.8   | 1.8   | 1.2   | 0.5   | 0.1   | 0.2  | 0.1   |
| 471K  | 7.5   | 5.0   | 4.2   | 1.3   | 0.9   | 0.4   | 0.2   | 0.1  | 0.1   |
| 424K  | 7.9   | 4.2   | 2.7   | 1.7   | 0.8   | 0.2   | 0.1   | 0.2  | 0.0   |
| 381K  | 8.4   | 5.4   | 2.1   | 1.3   | 0.3   | 0.3   | 0.1   | 0.1  | 0.0   |
| 343K  | 0.1   | 5.6   | 2.9   | 1.3   | 0.2   | 0.1   | 0.1   | 0.0  | 0.0   |
| 310K  | 8.2   | 3.7   | 1.6   | 1.6   | 0.2   | 0.0   | 0.0   | 0.0  | 0.0   |

Table 3. Excerpt of the percentage population matrix for the gonadotropin-releasing hormone at 310K.

| Bond | -170° | -160° | -150° | -140° | -130° | -120° | -110° | -100° | -90° | -80° | -70° |
|------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|------|
| 1    | 10.1  | 9.1   | 5.6   | 1.7   | 0.5   | 0.3   | 0.0   | 0.0   | 0.1  | 0.0  | 0.1  |
| 2    | 1.4   | 3.8   | 6.6   | 9.8   | 11.8  | 10.0  | 12.3  | 10.4  | 11.4 | 9.0  | 8.1  |
| 3    | 4.2   | 0.3   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0  | 0.1  | 0.0  |
| 4    | 29.4  | 31.9  | 5.2   | 0.8   | 0.0   | 0.0   | 0.0   | 0.0   | 0.1  | 0.2  | 1.1  |
| 5    | 0.3   | 1.1   | 1.0   | 1.9   | 3.5   | 3.9   | 5.1   | 7.9   | 19.0 | 19.3 | 9.4  |
| 6    | 0.5   | 2.7   | 4.7   | 4.6   | 11.7  | 6.8   | 8.7   | 6.2   | 14.1 | 14.1 | 13.3 |
| 7    | 5.7   | 3.0   | 0.5   | 0.2   | 1.3   | 0.7   | 0.0   | 0.0   | 0.0  | 0.0  | 0.0  |
| 8    | 21.2  | 14.1  | 7.4   | 3.6   | 0.0   | 0.0   | 0.0   | 0.0   | 0.2  | 0.4  | 1.7  |
| 9    | 0.0   | 0.0   | 0.1   | 0.6   | 0.6   | 0.1   | 0.3   | 0.3   | 5.9  | 2.0  | 3.3  |
| 10   | 0.0   | 2.3   | 10.9  | 19.5  | 16.0  | 12.9  | 5.6   | 2.3   | 6.1  | 8.9  | 5.8  |
| 11   | 1.4   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0  | 0.2  | 1.3  |
| 12   | 0.4   | 0.1   | 0.1   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.4  | 1.4  | 10.1 |
| 13   | 2.3   | 1.6   | 1.5   | 0.6   | 0.7   | 0.7   | 0.6   | 0.6   | 1.1  | 1.6  | 1.4  |
| 14   | 0.3   | 0.5   | 0.9   | 4.5   | 3.3   | 9.5   | 10.8  | 8.2   | 10.1 | 10.3 | 18.7 |

CLAIMS

1. A method for identifying structurally active molecule comprising
  - (a) performing multiple simulated annealing runs
  - 5 in order to reveal populated and unpopulated regions of multidimensional conformation space; and,
  - (b) performing a simulation at a fixed temperature, with sampling only from populated regions found in the first step.

1/27

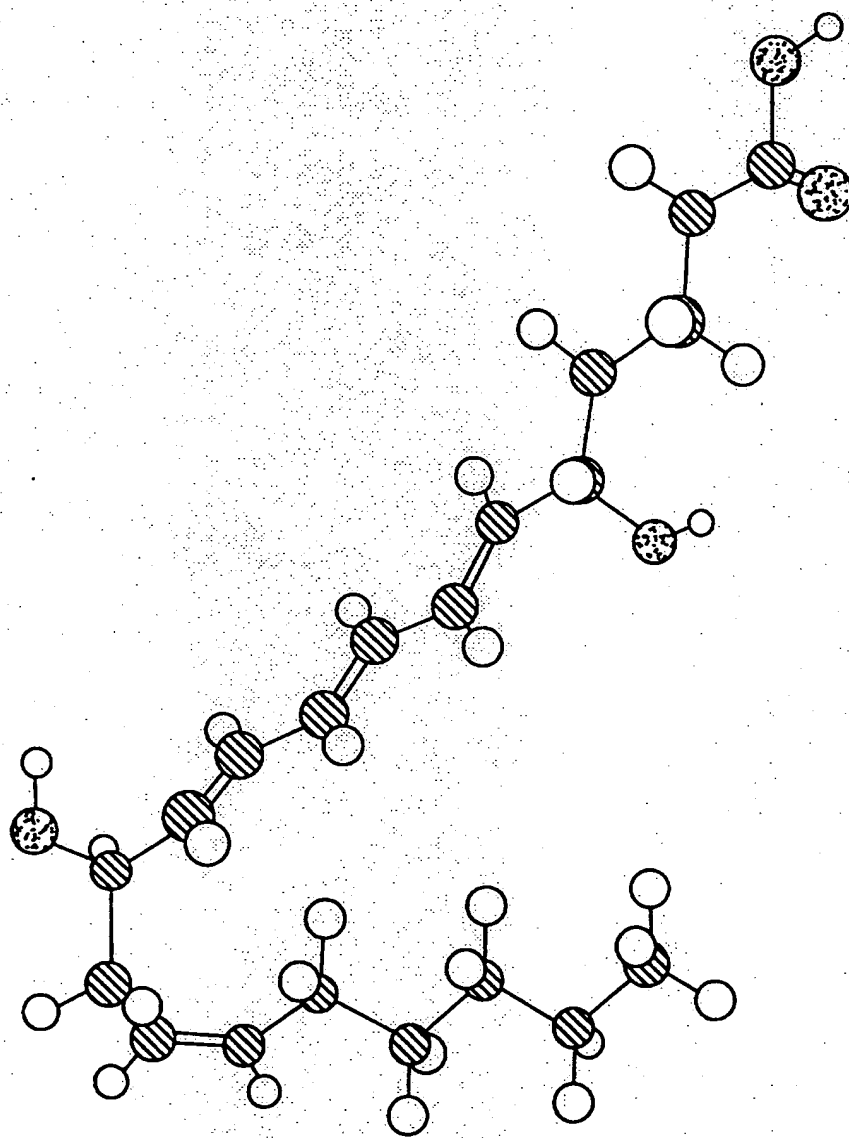


FIG. 1

2/27

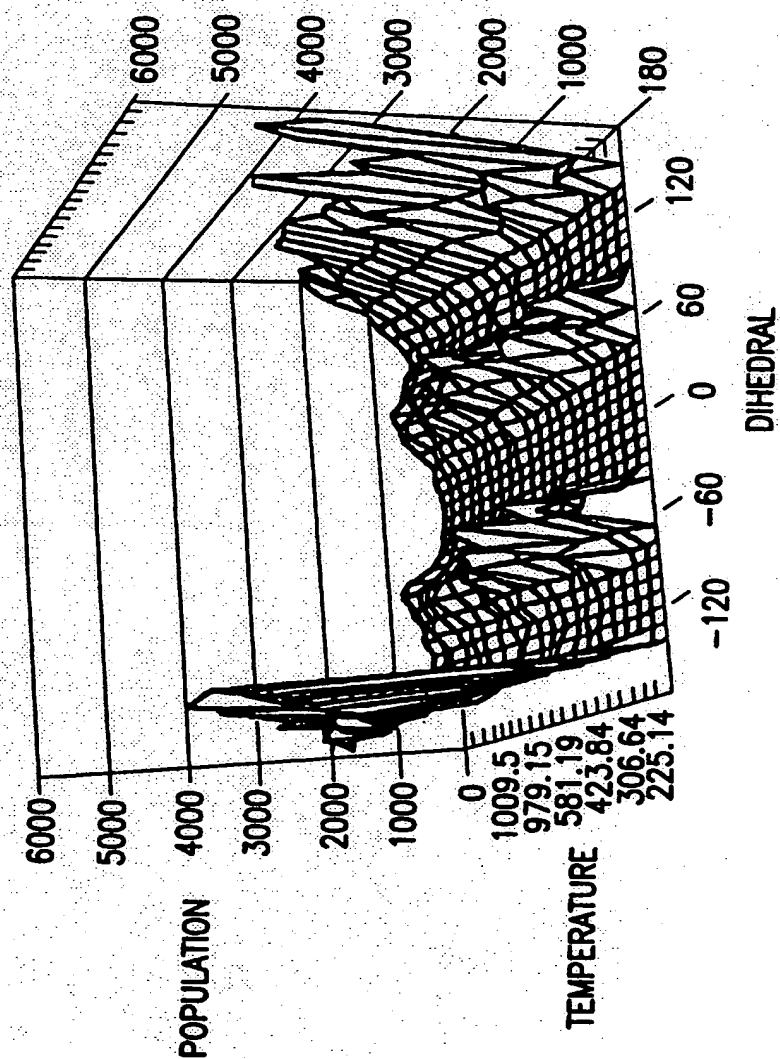


FIG.2



3/27

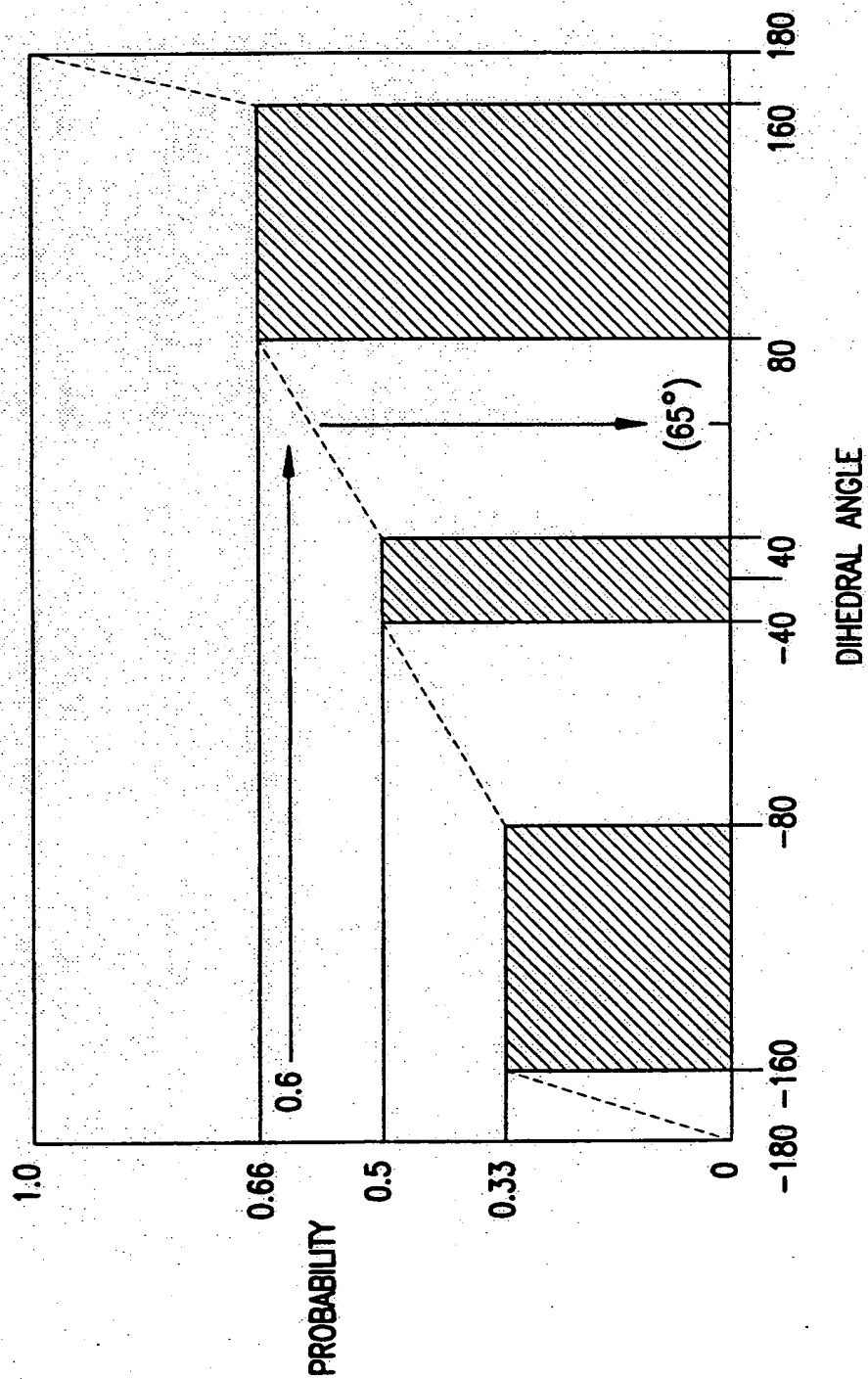


FIG. 3

4/27

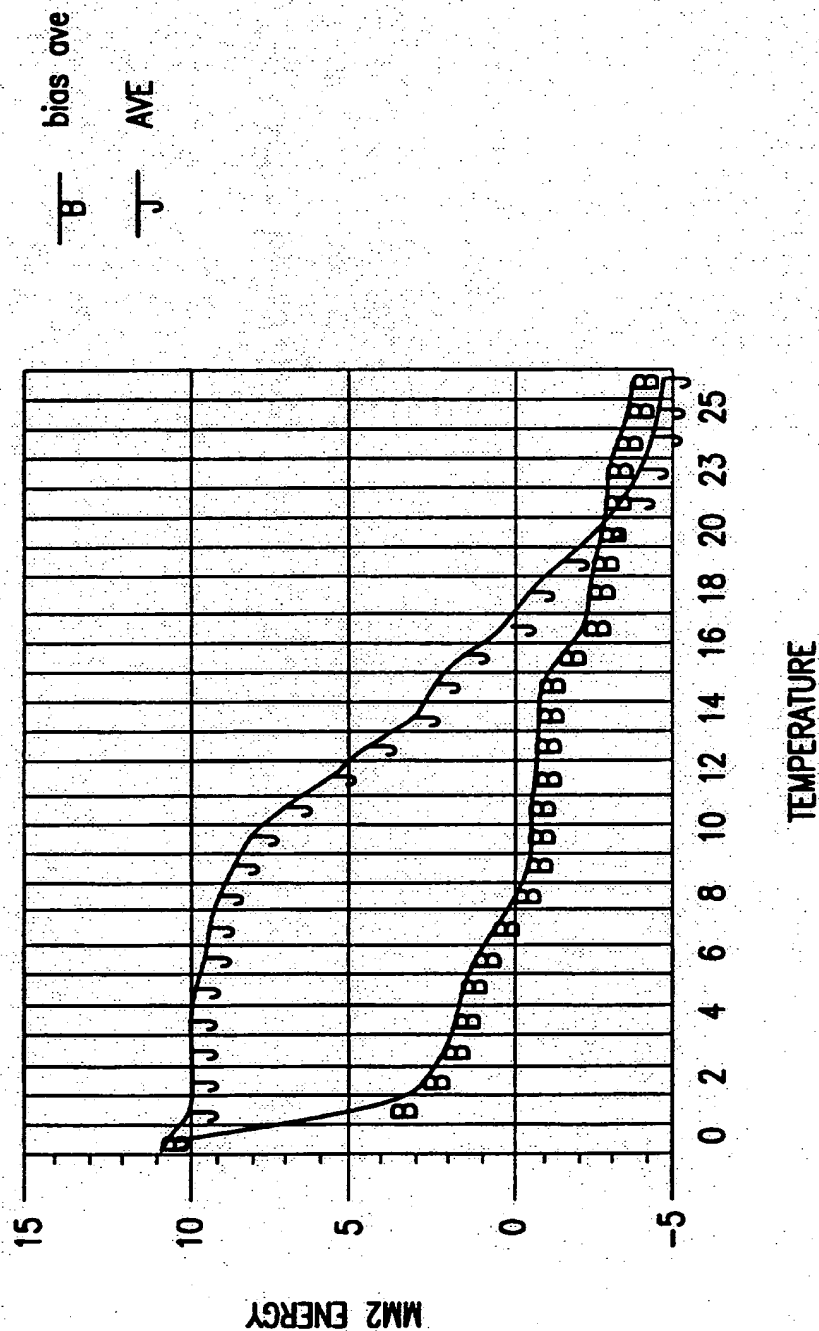


FIG.4

5/27

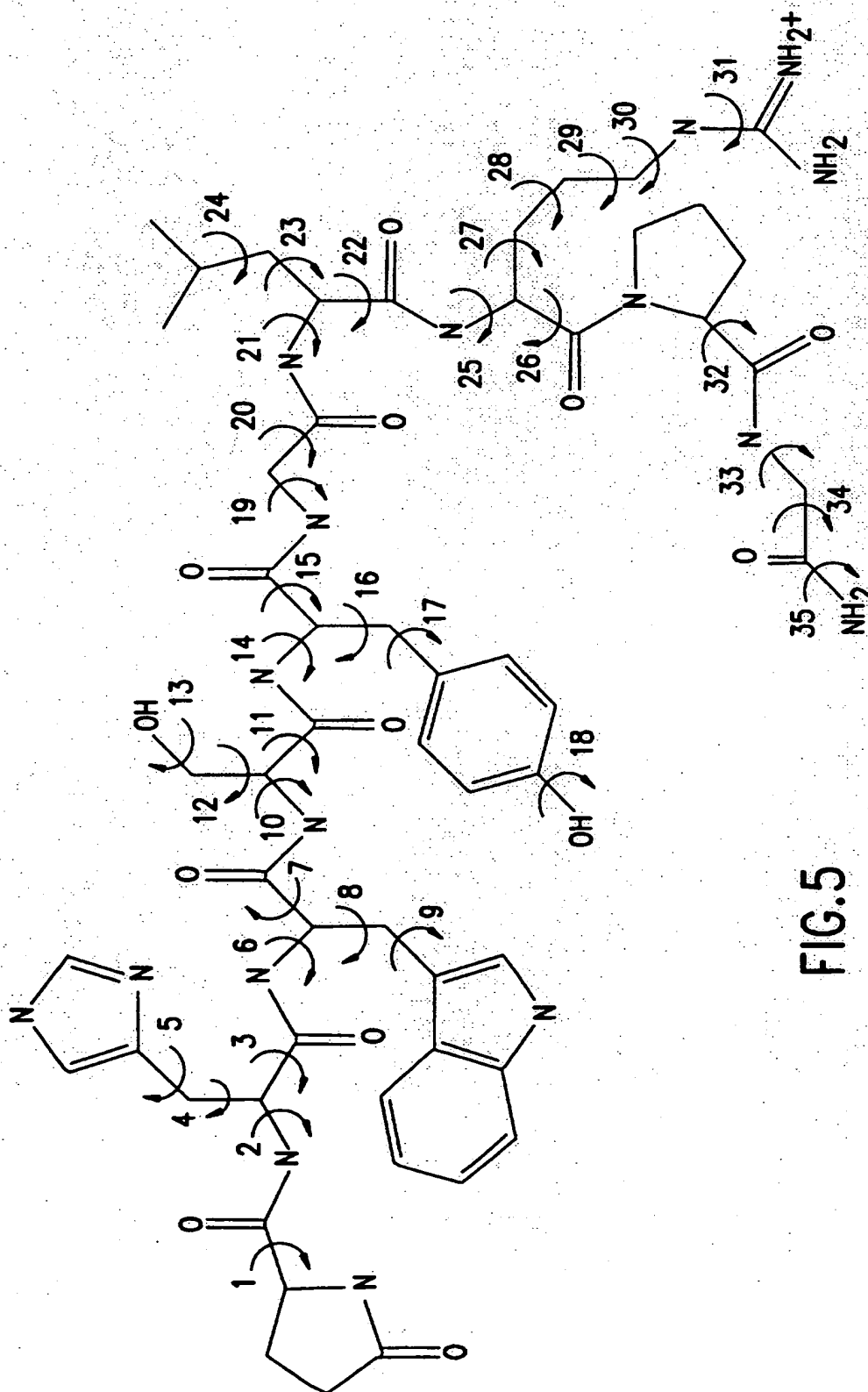


FIG.5

SUBSTITUTE SHEET (RULE 26)

6/27

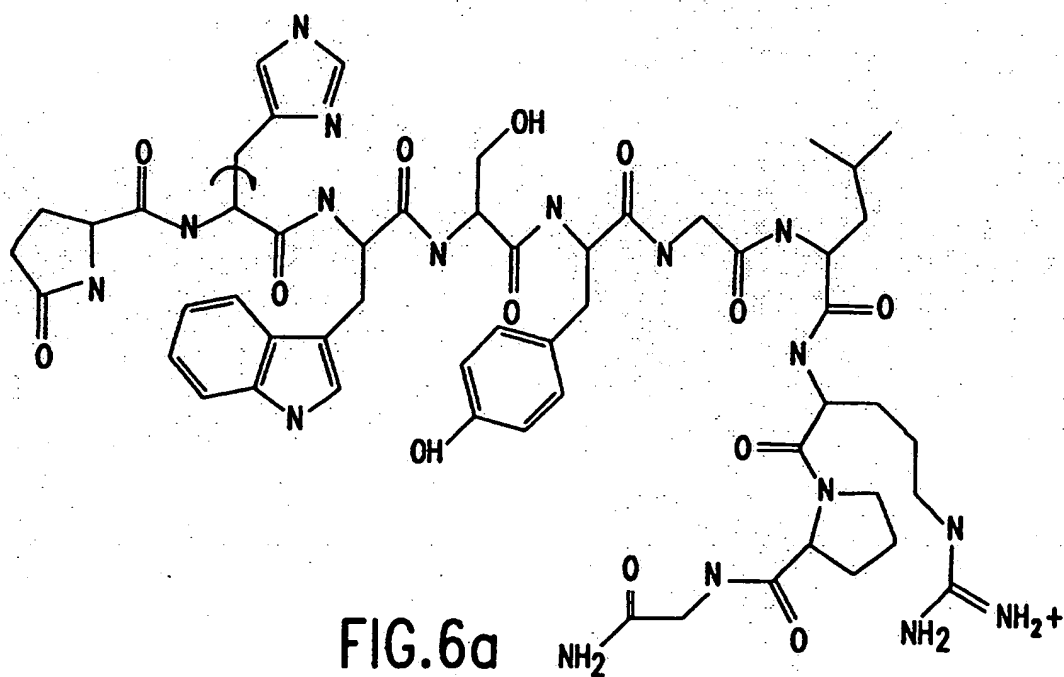
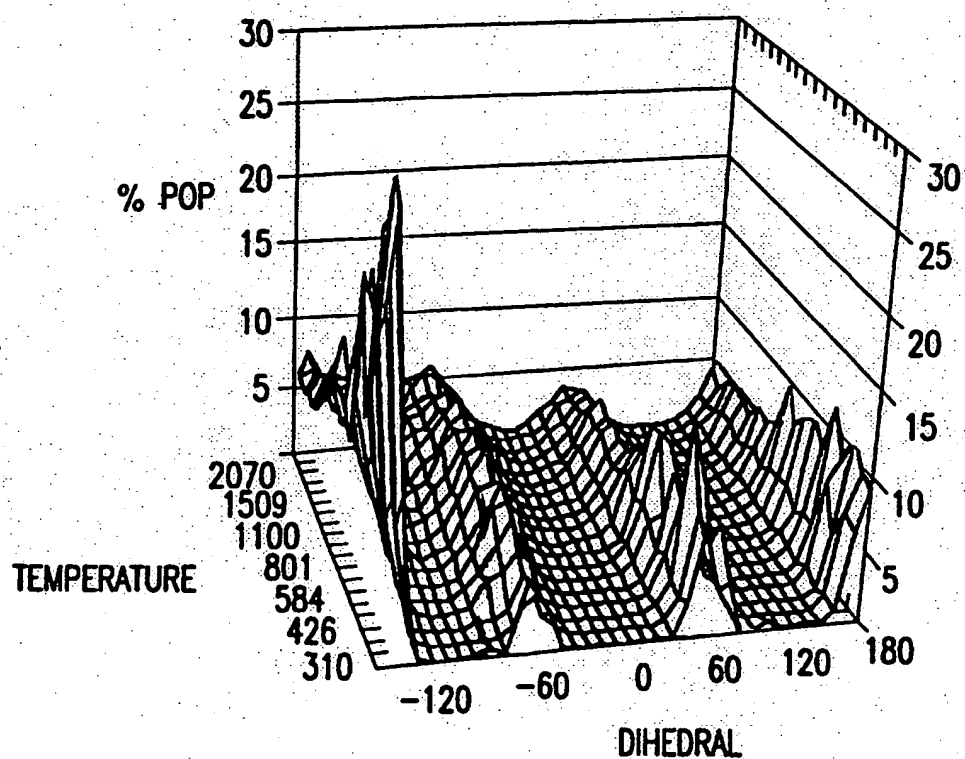
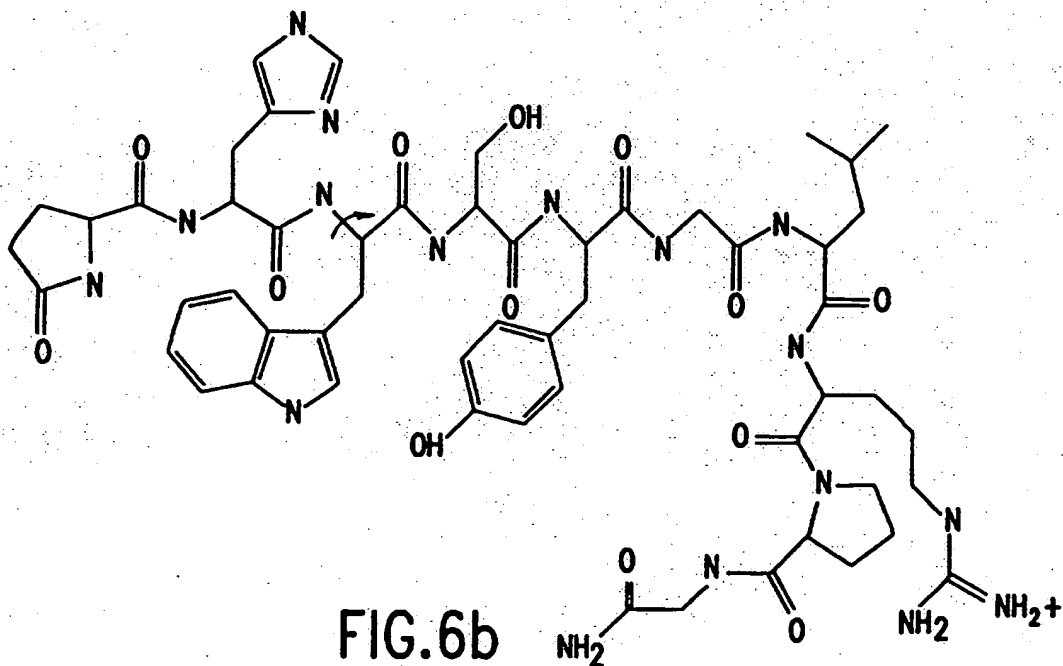
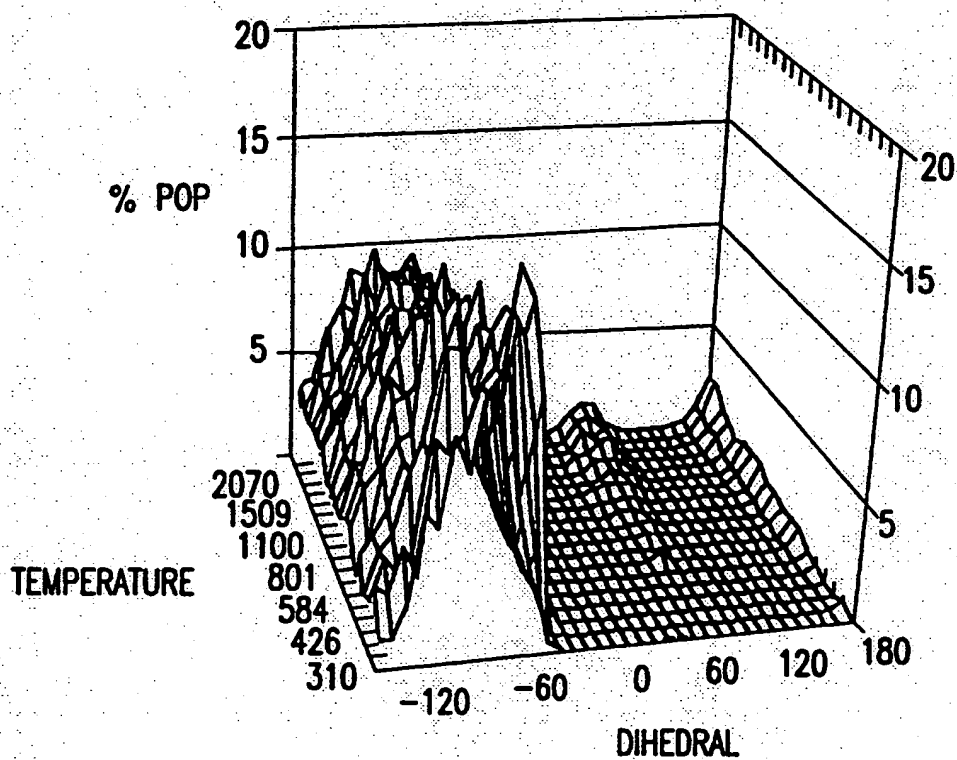


FIG. 6a

7/27



SUBSTITUTE SHEET (RULE 26)

8/27

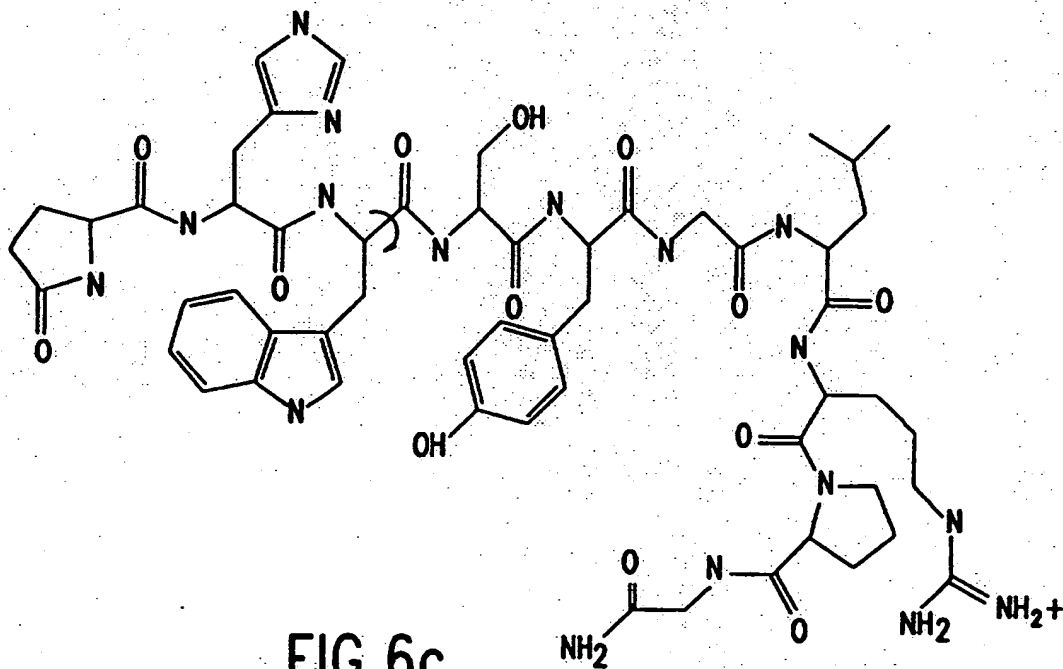
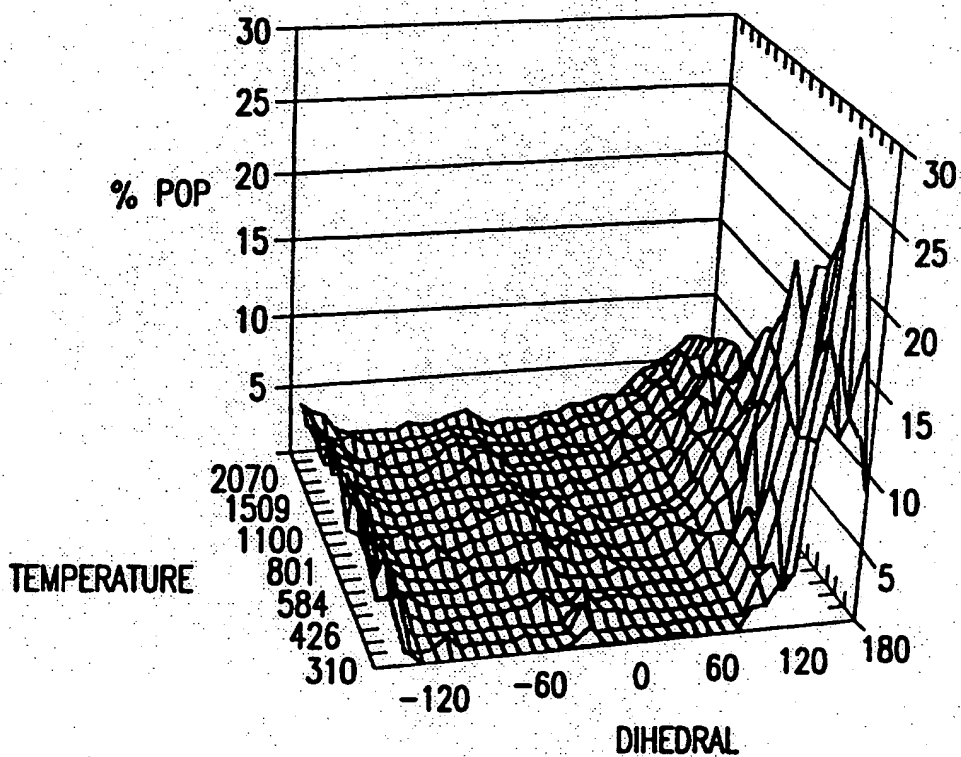


FIG. 6c

SUBSTITUTE SHEET (RULE 26)

9/27

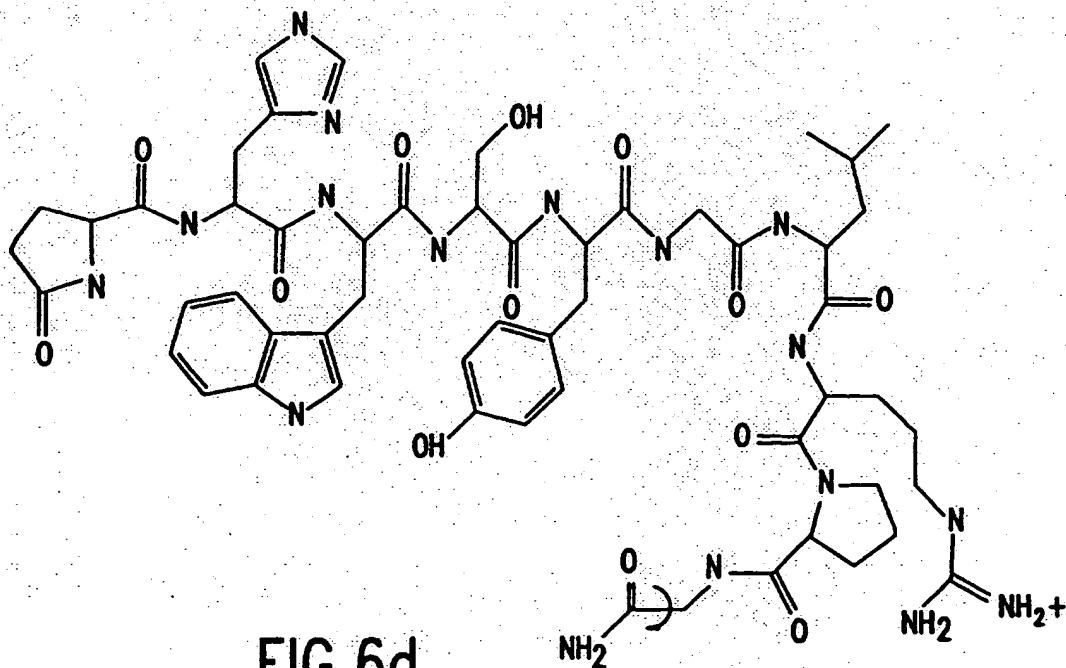
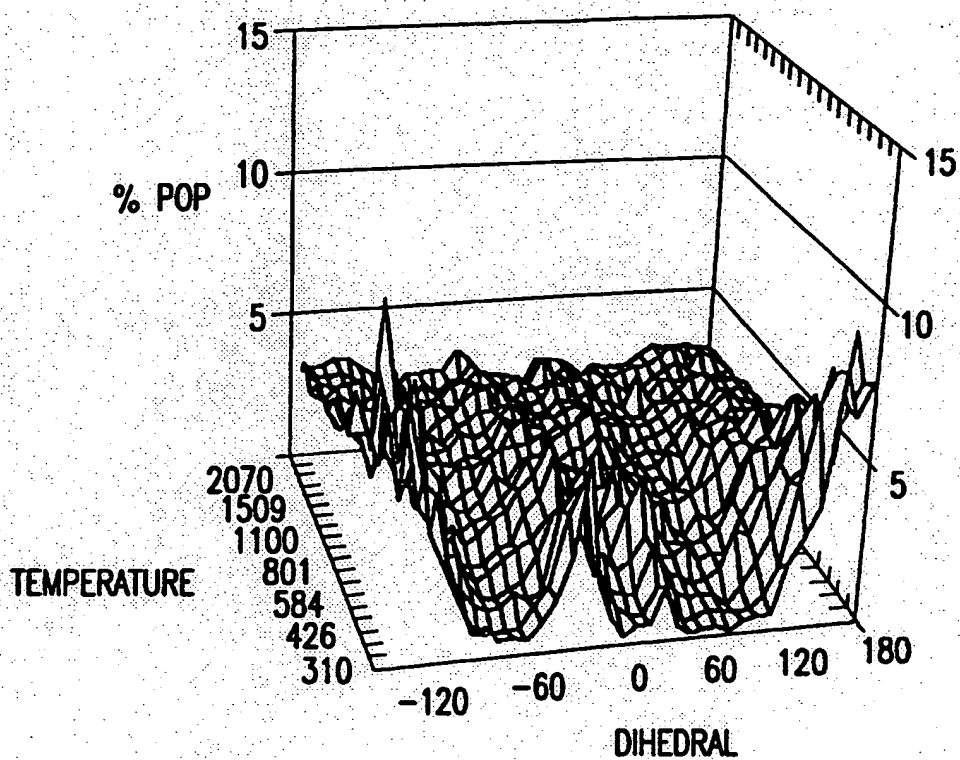


FIG.6d

SUBSTITUTE SHEET (RULE 26)

10/27

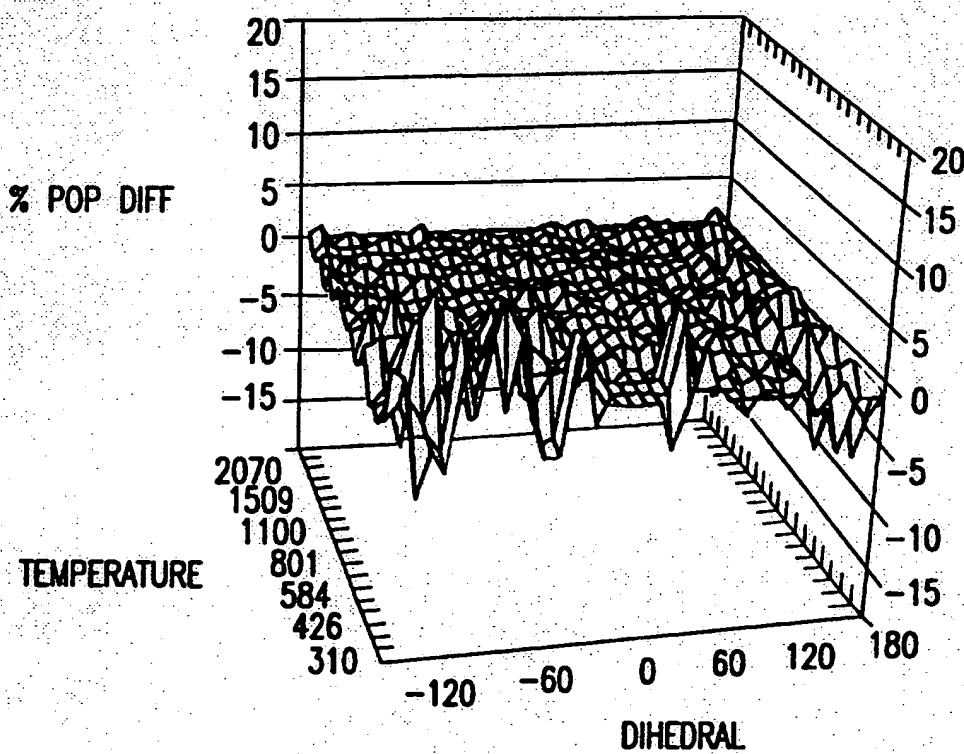


FIG.7a



11/27

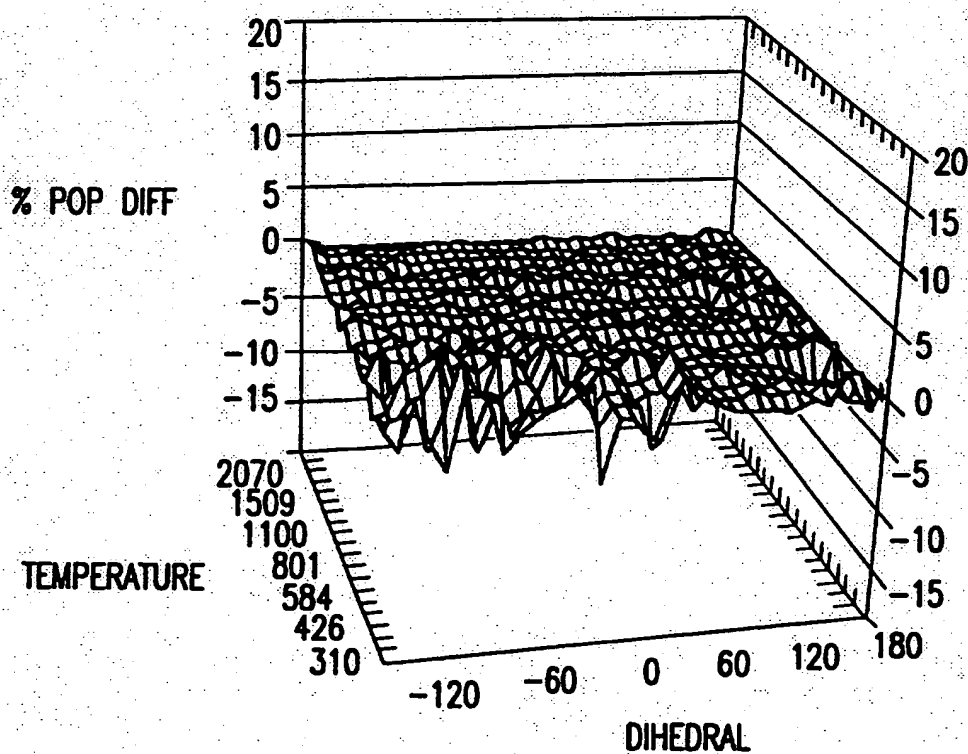


FIG.7b

12/27

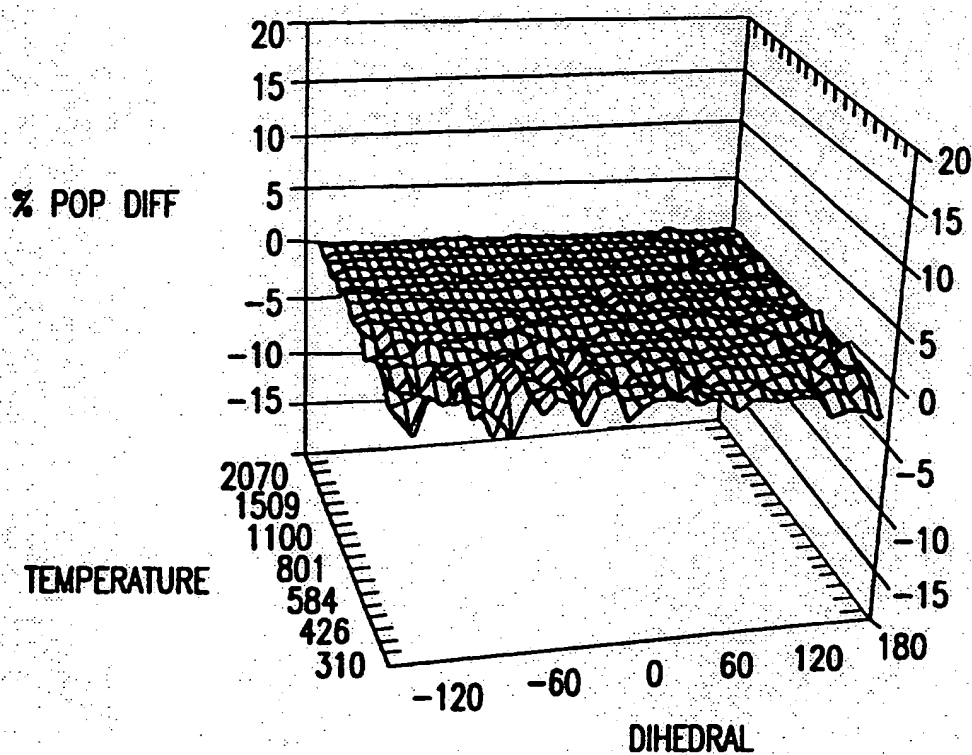


FIG.7c

13/27

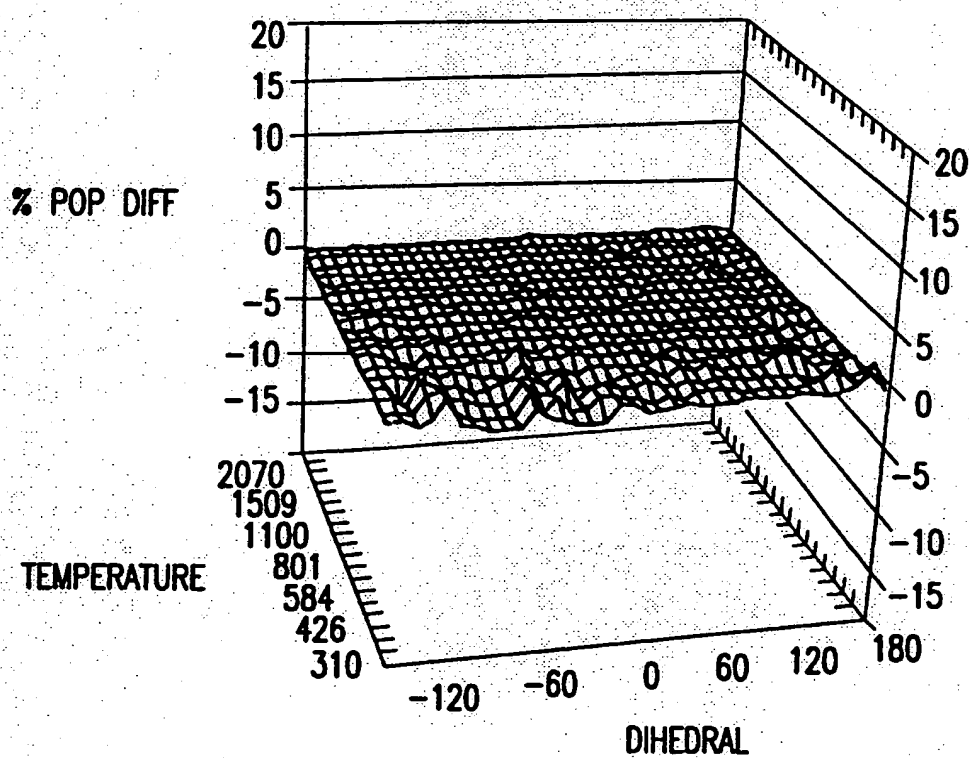


FIG.7d

14/27

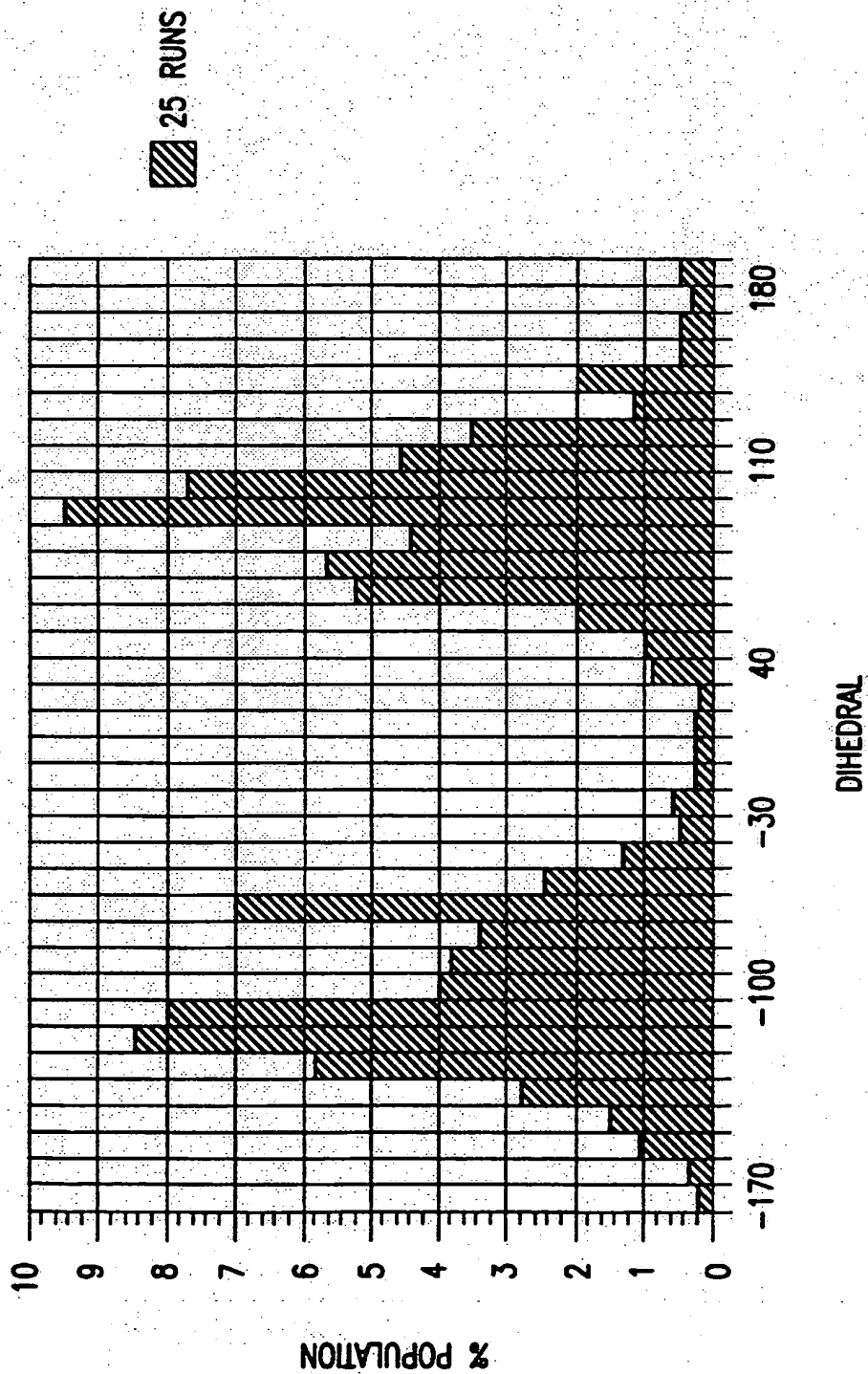


FIG.8a

15/27

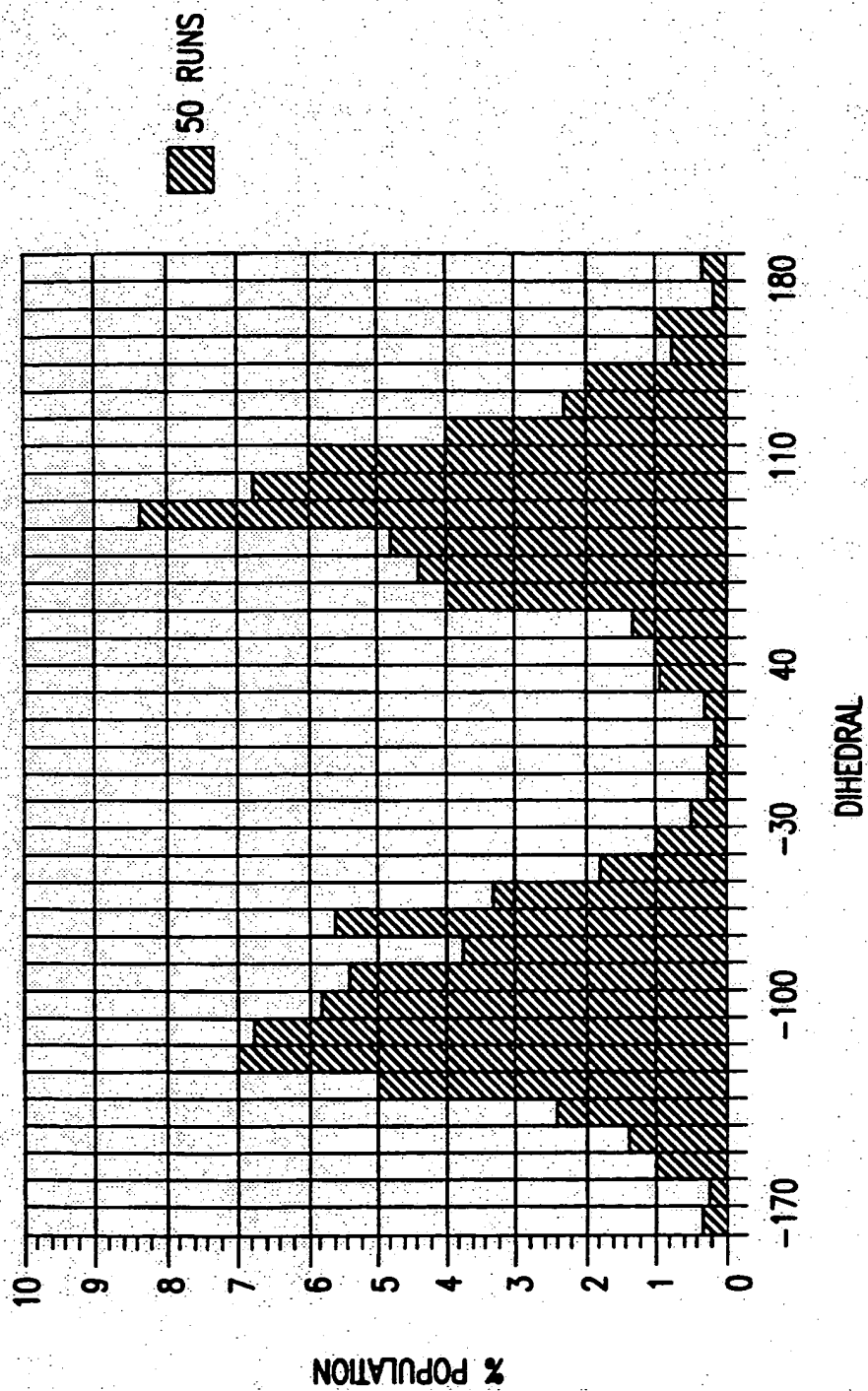


FIG.8b

16/27

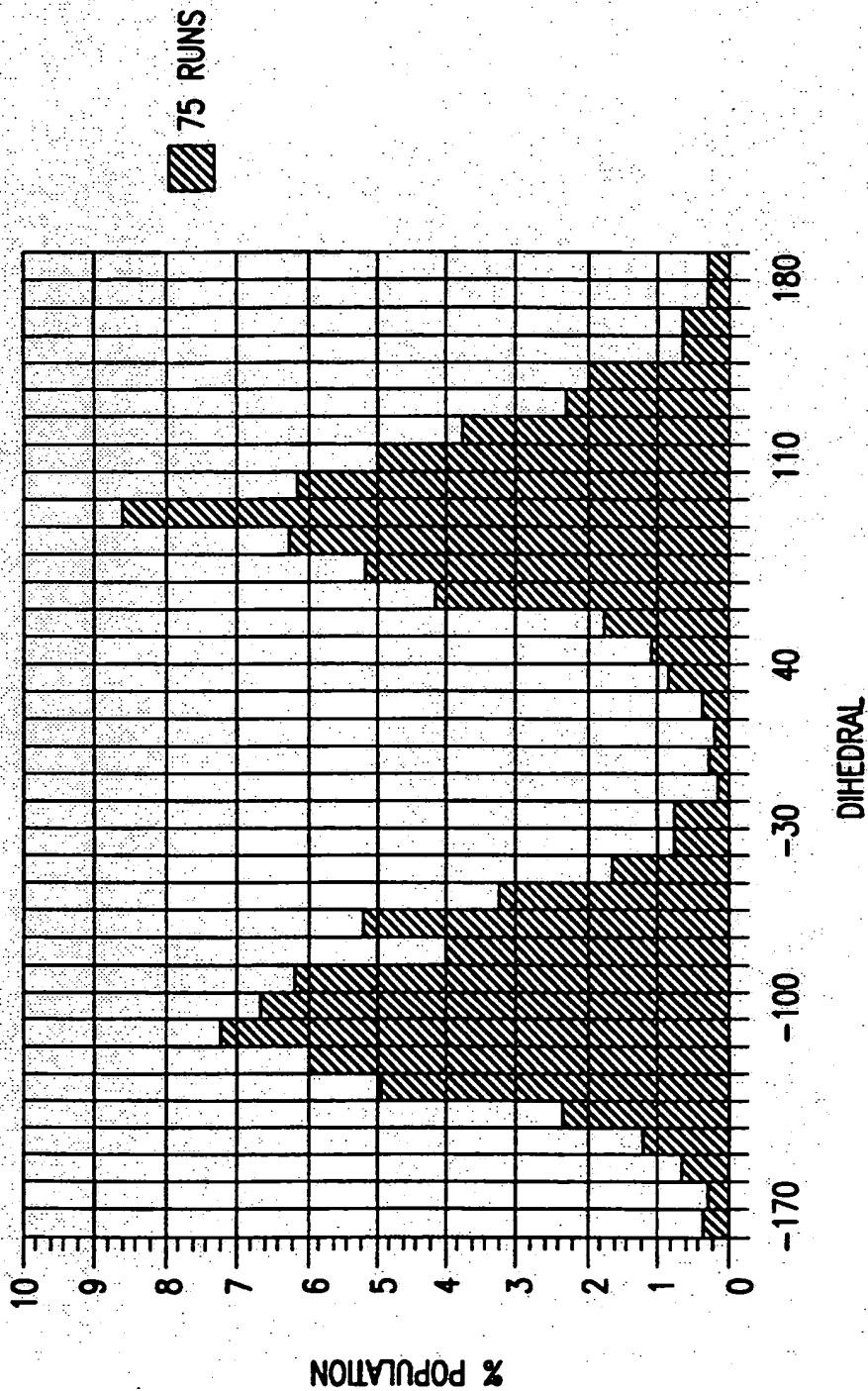


FIG. 8c

17/27

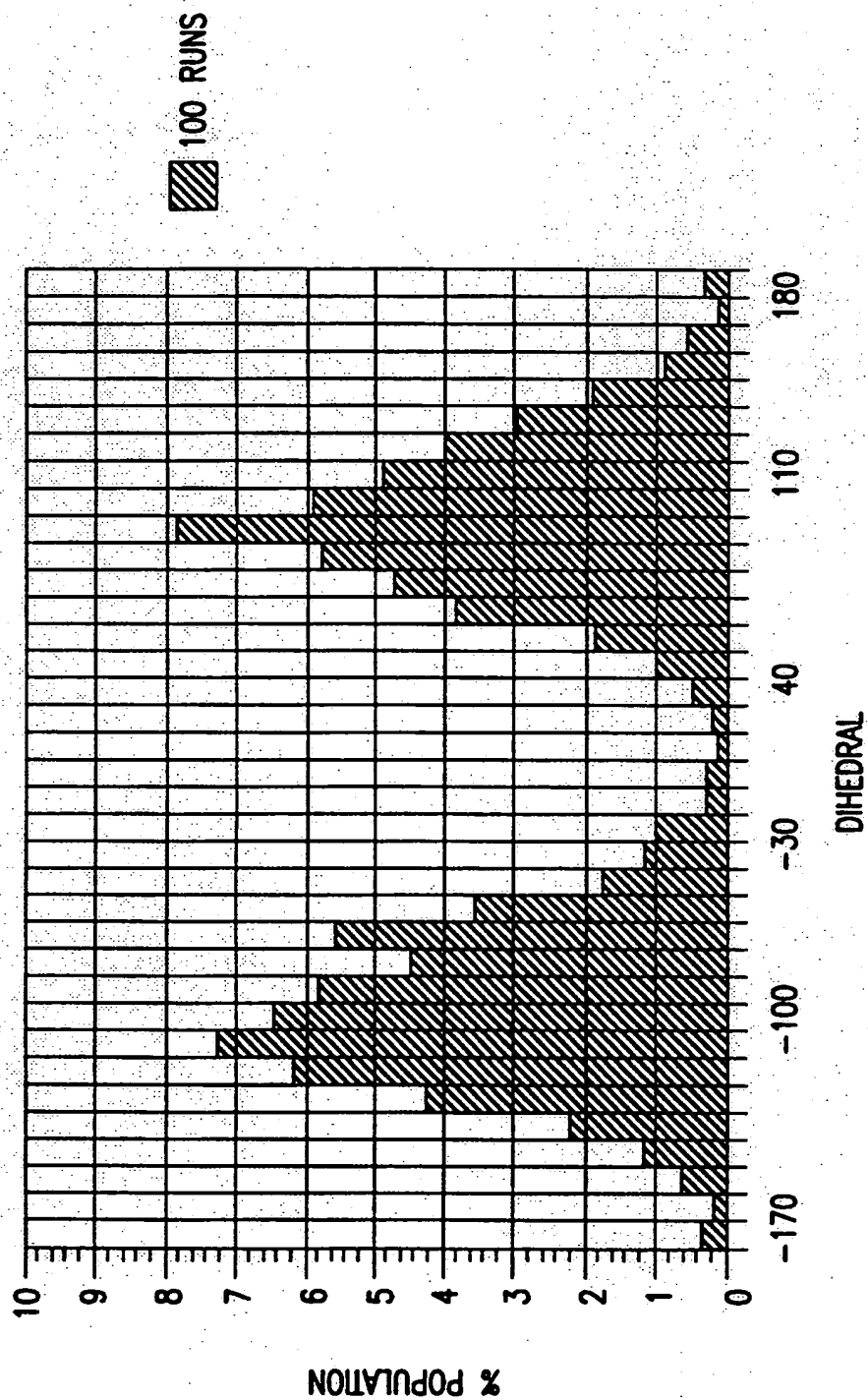


FIG.8d

18/27

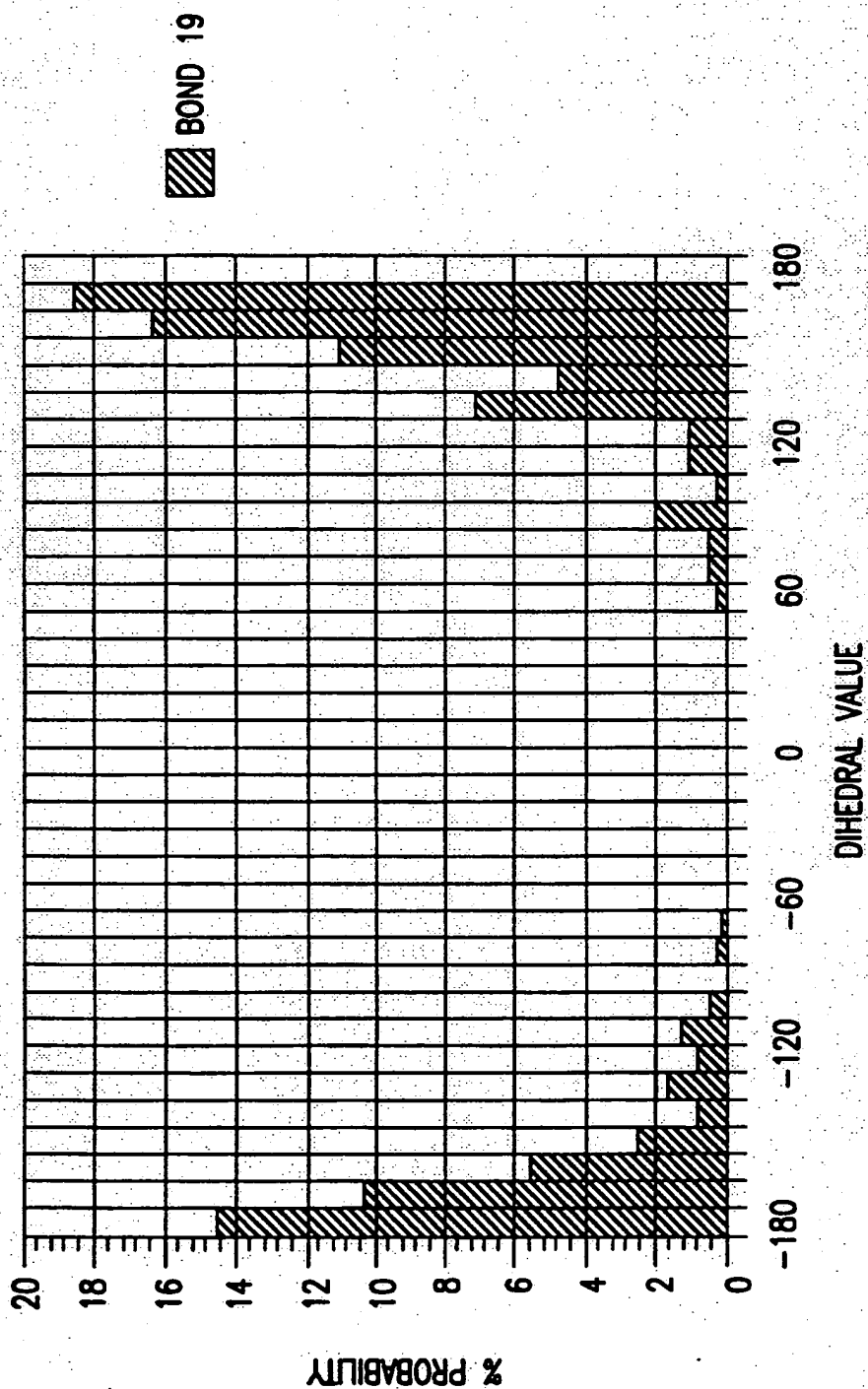


FIG.9a



19/27

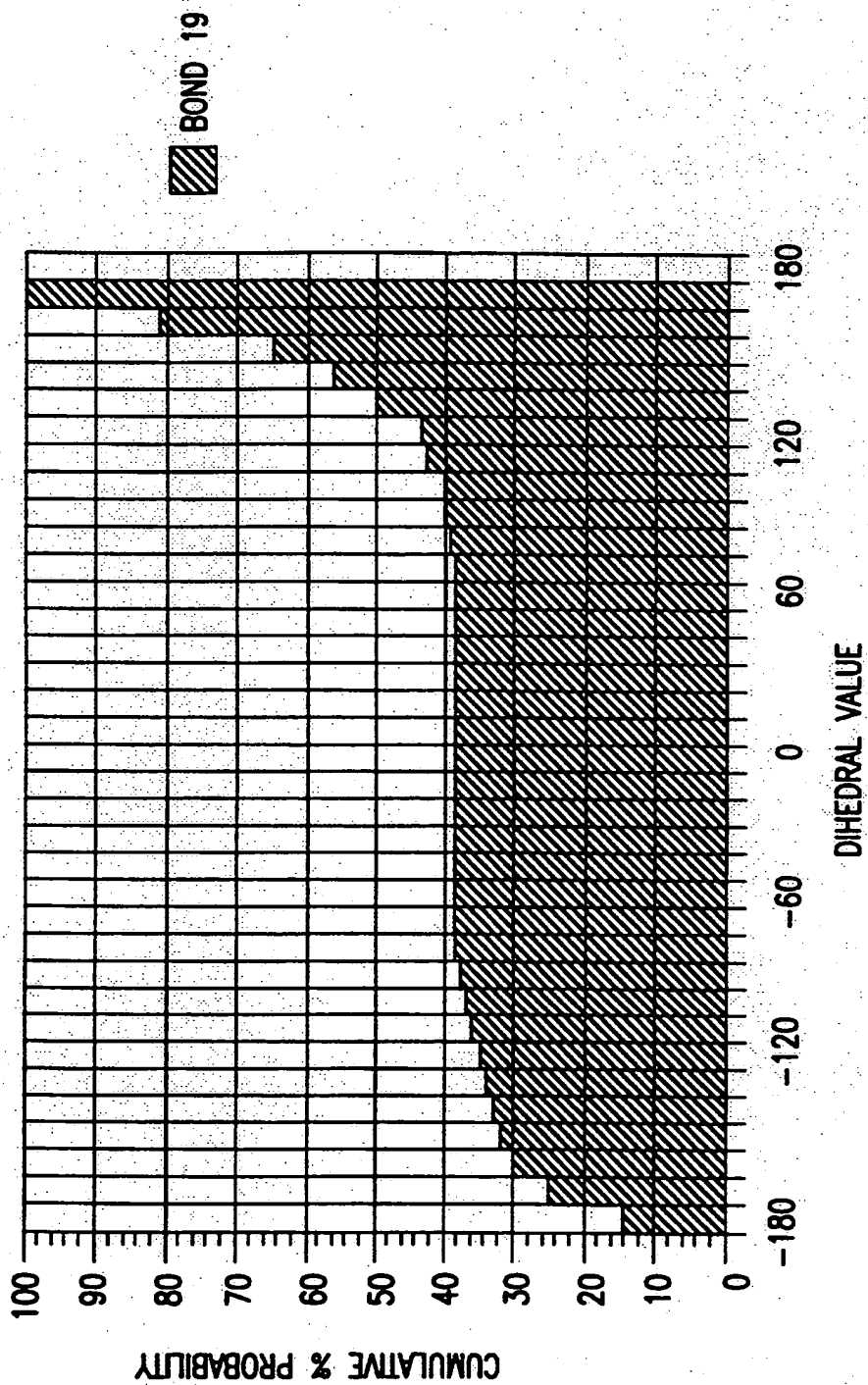


FIG.9b

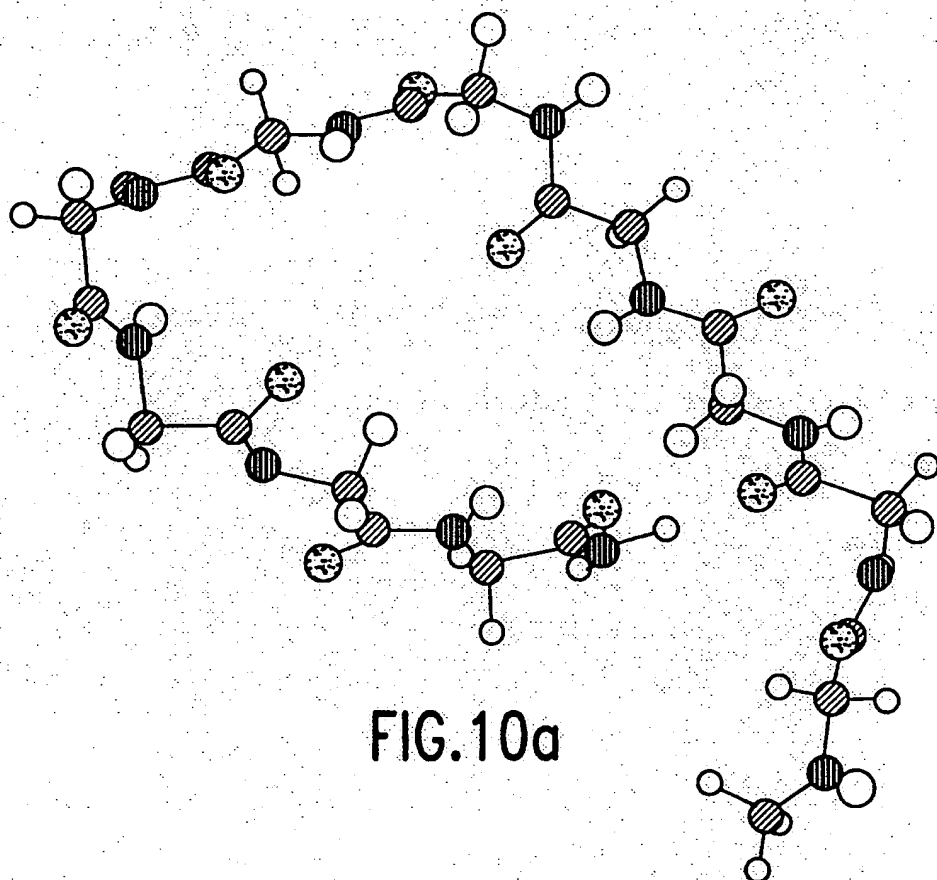


FIG.10a

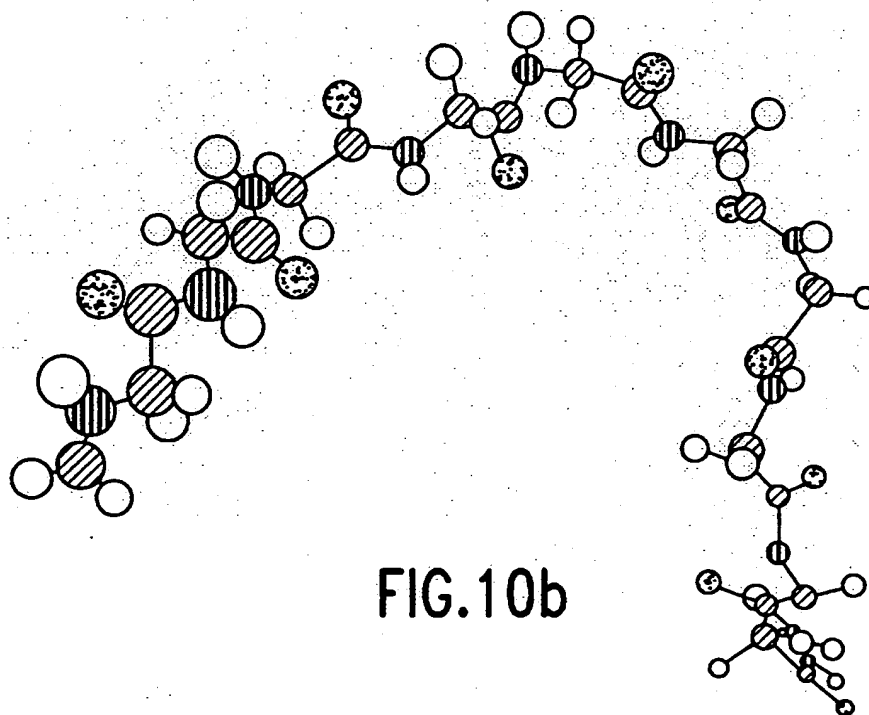


FIG.10b

21/27

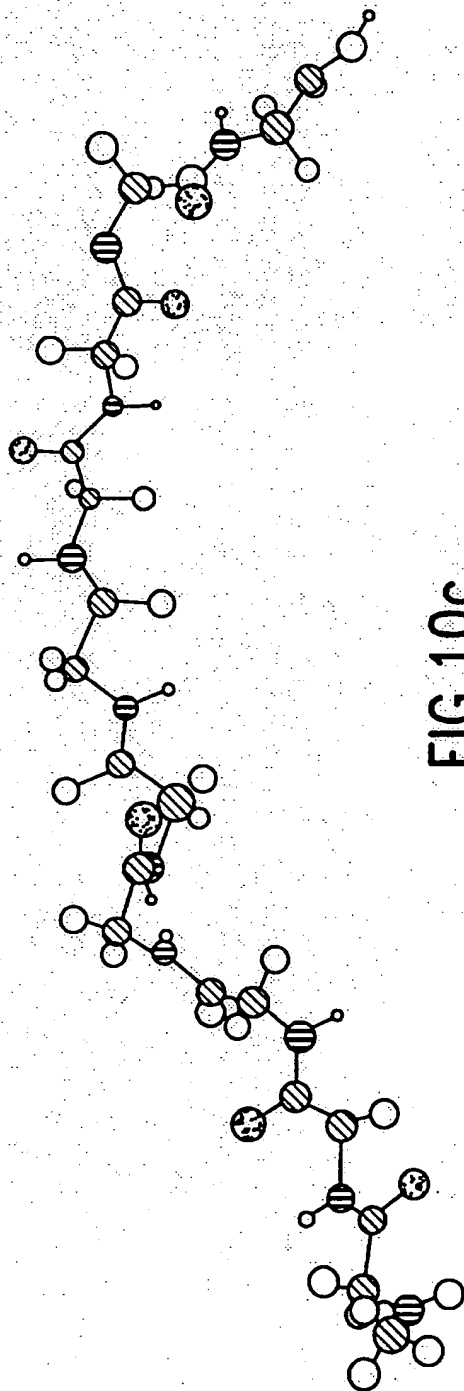
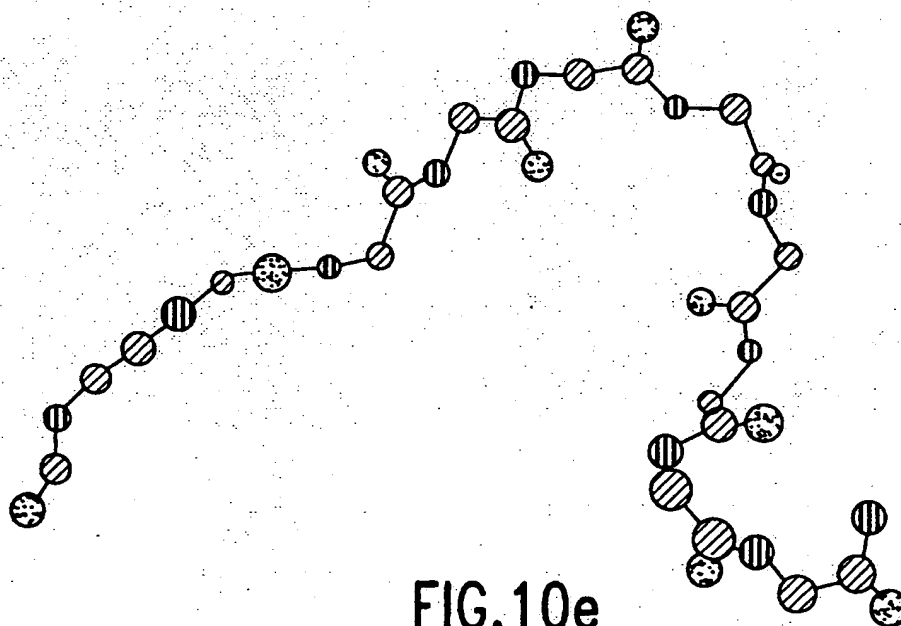
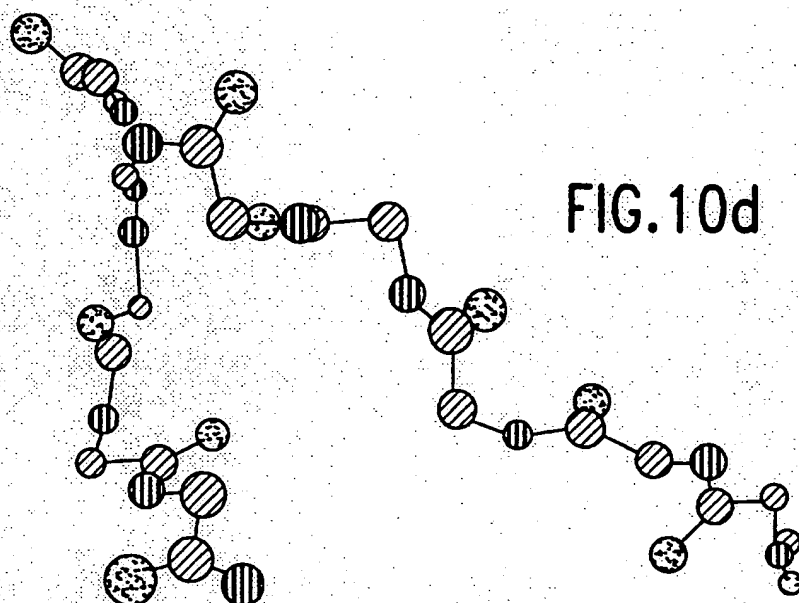


FIG.10c

22/27



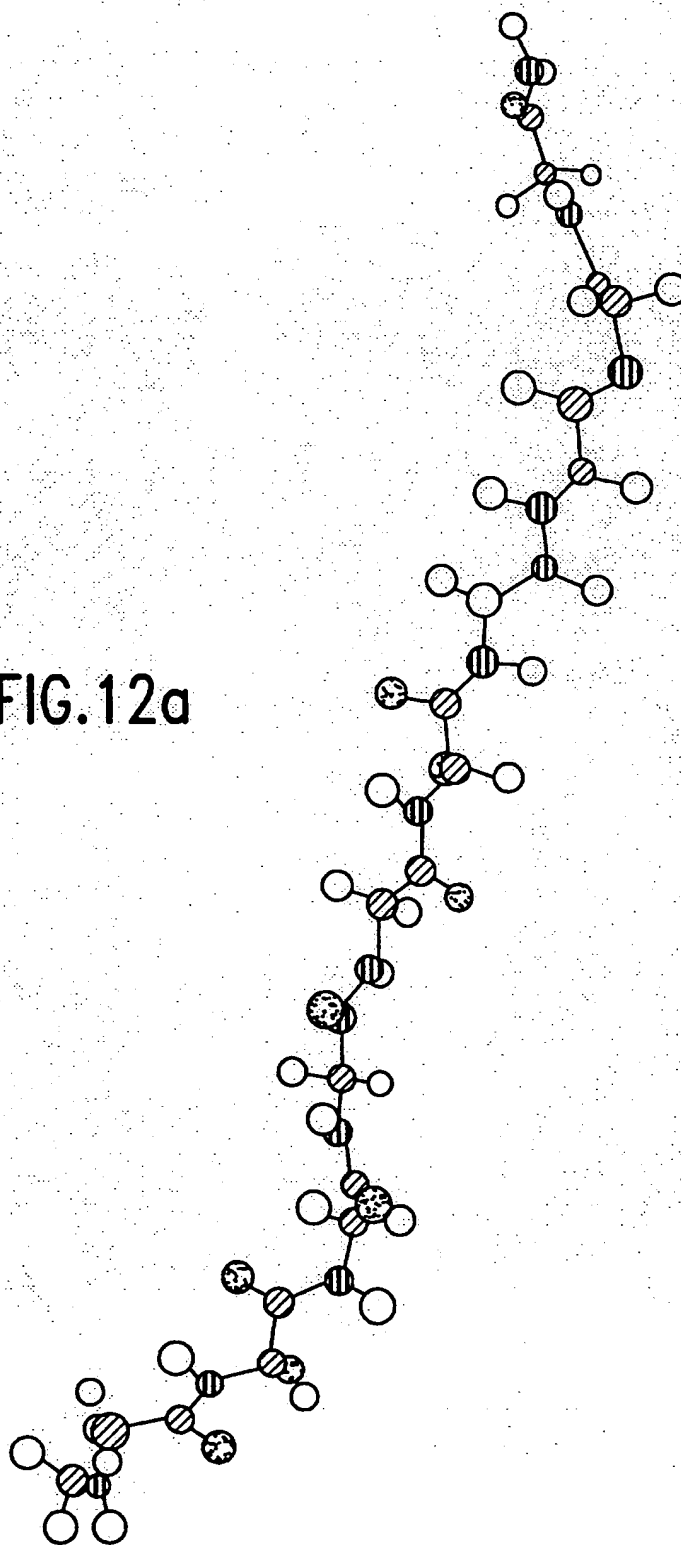
23 / 27



FIG. 11

24/27

FIG. 12a



25/27

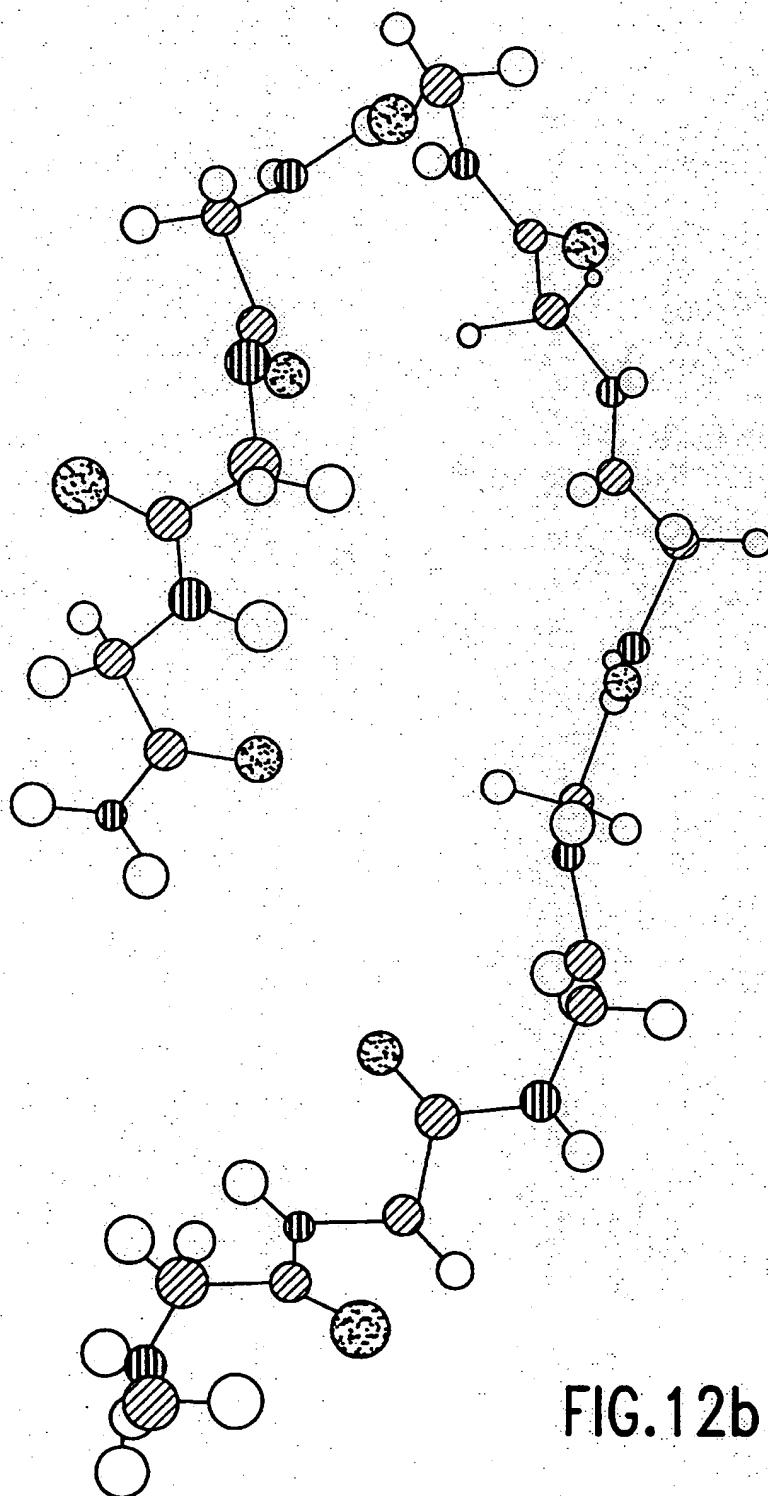


FIG.12b

26 / 27



FIG. 13



27/27

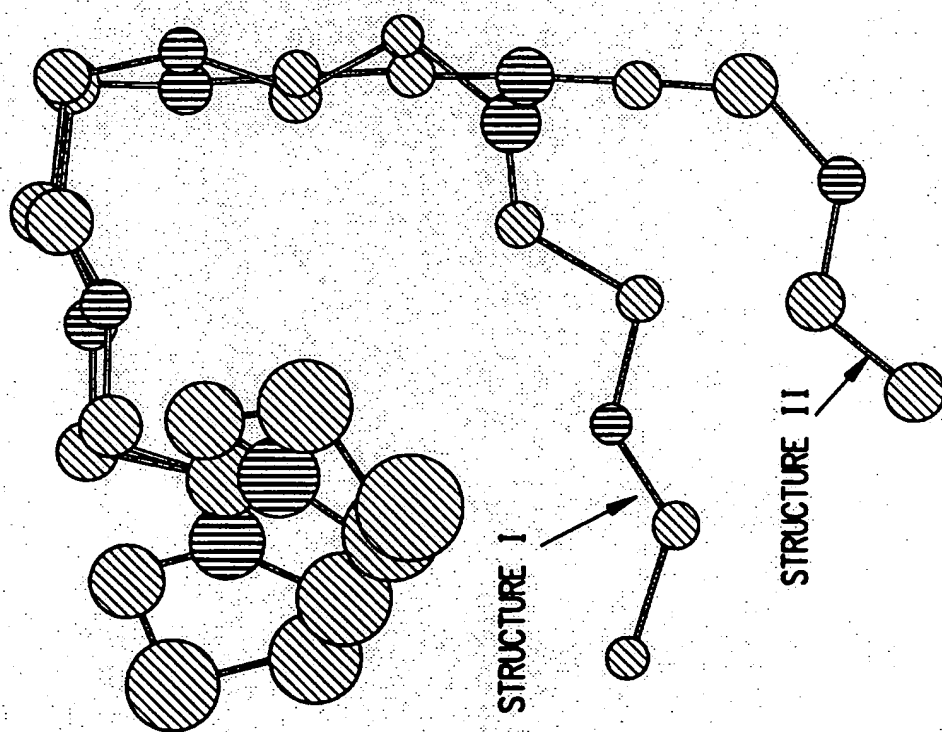


FIG.14

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US96/06110

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G06F 17/00, 159:00

US CL : 354/413.01

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 354/413.01, 497, 499, 578

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

STN, APS

search terms: simulated annealing, molecular dynamics, conformational searching, biased sampling

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No. |
|-----------|--|-----------------------|
| A         | WILSON et al. 'Conformational Analysis of Flexible Molecules.' In: Adaption of Simulated Annealing to chemical Optimization Problems. Edited by Kalivas. Elsevier Science B.V., 1995, pages 351-367, especially pages 359-366. | 1                     |
| X         | WILSON et al. Applications of Simulated Annealing to the Conformational Analysis of Flexible Molecules. Journal of Computational Chemistry. 1991, Vol. 12, No. 3, pages 342-349, see 346-347.                                  | 1                     |



Further documents are listed in the continuation of Box C.



See patent family annex.

|   |     |  |
|---|-----|--|
| * Special categories of cited documents:  | *T  | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  |
| *A* document defining the general state of the art which is not considered to be of particular relevance  | *X* | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone   |
| *E* earlier document published on or after the international filing date  | *Y* | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | *G* | document member of the same patent family  |
| *O* document referring to an oral disclosure, use, exhibition or other means  |     |  |
| *P* document published prior to the international filing date but later than the priority date claimed  |     |  |

Date of the actual completion of the international search

26 JULY 1996

Date of mailing of the international search report

22 AUG 1996

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

LORA M. GREEN

Telephone No. (703) 308-0196